Investigations on the Influence of Acoustics on Live Music Performance using Virtual Acoustic Methods

Sebastià Vicenç Amengual Garí

Ph.D. Dissertation

Investigations on the Influence of Acoustics on Live Music Performance using Virtual Acoustic Methods

Sebastià Vicenç Amengual Garí

Erich Thienhaus Institut

1. Reviewer Prof. Dr.-Ing. Malte Kob
Erich-Thienhaus-Institut
Detmold University of Music

2. Reviewer Prof. Dr.-Ing. Dr. Tech. h. c. Jens Blauert
Institut für Kommunikationsakustik
Ruhr-Universität Bochum

1. Supervisor Prof. Dr.-Ing. Malte Kob
Erich Thienhaus Institute
Detmold University of Music

2. Supervisor Prof. Tapio Lokki, D.Sc.(Tech.)
Department of Computer Science
Aalto University

September 2017

Sebastià Vicenç Amengual Garí
Investigations on the Influence of Acoustics on Live Music Performance using Virtual Acoustic Methods
Ph.D. Dissertation, September 2017
Reviewers: Prof. Dr.-Ing. Malte Kob and Prof. Dr.-Ing. Dr. Tech. h. c. Jens Blauert
Supervisors: Prof. Dr.-Ing. Malte Kob and Prof. Tapio Lokki, D.Sc.(Tech.)

Detmold University of Music
Erich Thienhaus Institut
Neustadt 22
32756 Detmold
Germany

Abstract

Room acoustic conditions are an inherent element of every live music performance. They interact with the sound that is generated by the musicians, modifying the characteristics of the sound received by audience and musicians. While listeners usually play a passive role in the context of a live performance, musicians are part of a feedback loop composed by themselves, their instruments, and the room. The goal of this thesis is to characterize the effects of room acoustics in live performances, by studying the acoustical preferences of musicians and characterizing potential performance adjustments implemented by solo players while adapting their interpretation to the room acoustic conditions.

To conduct systematic experiments, a virtual acoustic environment that replicates acoustic conditions of real rooms in laboratory conditions is implemented. Room impulse responses of performance rooms are measured and parametrized using spatial measurement techniques. The responses are later resynthesized and convolved in real-time with the sound generated by a musician. The resulting sound is reproduced through a 3D loudspeaker set-up, allowing musicians to perform under replicated acoustic conditions of measured rooms in real-time. The system is used to conduct pilot studies on stage acoustics preferences of semi-professional trumpet players, and to study the impact of room acoustics on potential performance adjustments of live performance. To this end, musical pieces are recorded under different acoustic conditions and later analyzed. A second experiment is performed with organ players in the Detmold Konzerthaus. The reverberation time of the hall is modified using a reverberation enhancement system, and live performances are recorded under different acoustic conditions using a MIDI interface. Similarly to the trumpet players, the recordings are analyzed to evaluate the extent of the performance adjustments. Finally, listening tests are conducted to assess the perceived impact of those adjustments by listeners.

Results of the experiments suggest that musicians systematically adjust their performance to accommodate room acoustic conditions and listeners are generally able to perceive these changes. Trumpet players tend to decrease the sound level and sound brightness when exposed to longer and stronger reverberation. Some players adjust as well musical dynamics and aspects related to the *tempo* of their performance, although generalized trends are not observed. Dry environments are usually preferred to practice instrument technique, while longer reverberation times are preferred in concert conditions. Additionally, the presence of a sufficient amount of early energy contributes positively to the musicians' comfort, regardless of the direction of incidence of this sound energy. Organ players are prone to modifying the temporal aspects of the performance, generally decreasing the overall *tempo* and increasing the length of breaks in more reverberant environments. The musical character of the played excerpts seems to play an important role, and while for some pieces changes are generalized

and systematic, the performance of other pieces with soft dynamics and little contrast is generally less affected by room acoustics.

Zusammenfassung

Raumakustische Bedingungen sind ein inhärentes Element jeder Live-Musik-Darbietung. Sie interagieren mit dem Schall, der von den Musikern erzeugt wird, und modifizieren die Eigenschaften des Klangs, der von Publikum und Musikern empfangen wird. Während Hörer in der Regel eine passive Rolle im Kontext einer Live-Performance spielen, sind Musiker Teil einer Rückkopplungsschleife, die von ihnen selbst, ihren Instrumenten und dem Raum zusammengesetzt ist. Ziel dieser Arbeit ist es, die Effekte der Raumakustik in Live-Auftritten zu charakterisieren, indem die akustischen Präferenzen von Musikern studiert und potenzielle Leistungsanpassungen von Solo-Spielern charakterisiert werden während diese ihre Interpretation an die raumakustischen Bedingungen anpassen.

Um systematische Experimente durchzuführen zu können, wird eine virtuelle akustische Umgebung implementiert, indem die akustischen Bedingungen realer Räume unter Laborbedingungen repliziert werden. Raumimpulsantworten von Aufführungsräumen werden mit einem Mikrofon-Array gemessen und analysiert. Die Antworten werden später neu synthetisiert und in Echtzeit mit dem von einem Musiker erzeugten Klang gefaltet. Der resultierende Klang wird von einer 3D-Lautsprecheranordnung wiedergegeben, so dass Musiker unter replizierten akustischen Bedingungen der gemessenen Räume in Echtzeit spielen können. Das System wird verwendet, um Pilotstudien zur Bühnenakustik-Präferenz semiprofessioneller Trompetenspieler durchzuführen und die Auswirkungen der Raumakustik auf potenzielle Anpassungen des Spiels bei Live-Darbietungen zu untersuchen. Zu diesem Zweck werden Musikstücke unter verschiedenen akustischen Bedingungen aufgezeichnet und später automatisch analysiert. Ein zweites Experiment mit Orgelmusik wird im Konzerthaus Detmold durchgeführt. Die Nachhallzeit des Konzerthauses wird mit dem Raumakustik-Sytem *Vivace* modifiziert, und Live-Aufführungen werden unter verschiedenen akustischen Bedingungen mit einer MIDI-Schnittstelle aufgezeichnet. Ähnlich wie bei den Trompetenspielern werden die Aufnahmen analysiert, um das Ausmaß der Anpassungen des Spiels zu bewerten. Schließlich werden Hörversuche durchgeführt, um die wahrgenommenen Auswirkungen dieser Anpassungen auf die Zuhörer zu beurteilen.

Die Ergebnisse der Experimente deuten darauf hin, dass Musiker systematisch ihre Spielweise anpassen, um raumakustische Bedingungen zu berücksichtigen. Zuhörer können diese Veränderungen in der Regel wahrnehmen. Trompetenspieler neigen dazu, den Schallpegel und die Klanghelligkeit zu verringern, wenn sie in längerem und stärkerem Nachhall ausgesetzt sind. Einige Spieler passen auch dynamische und zeitliche Aspekte ihres Spiels an. Trockene Umgebungen werden in der Regel bevorzugt, um die Spieltechnik zu verbessern, während längere Nachhallzeiten bei Konzerten bevorzugt werden. Darüber hinaus trägt

das Vorhandensein einer ausreichenden Menge an frühe Schallenergie positiv zum Befinden der Musiker bei, unabhängig von der Einfallsrichtung dieser Schallanteile. Organisten sind empfindlich für zeitliche Änderungen ihres Spiel; in der Regel senken Sie das Tempo und erhöhen die Pausenlänge in Umgebungen mit längerem Nachhall. Der musikalische Charakter der gespiegelten Musikstücke scheint eine wichtige Rolle zu spielen: während für einige Stücke Änderungen einheitlich und systematisch sind, ist die Spielweise anderer Stücke mit geringerer Dynamik und weniger Kontrast in der Regel weniger von der Raumakustik betroffen.

In dieser Arbeit wird die männliched Form vertretend für weibliche und männliche Bezeichnungen verwendet.

Acknowledgements

First and foremost, I would like to thank my supervisors, Prof. Dr. Malte Kob and Prof. Dr. Tapio Lokki. I want to thank Prof. Kob for his confidence in my work since the very first day, and for the freedom and support that I had during these years to plan my ideas and bring them to reality without restrictions. Thanks to Prof. Lokki for being always ready to discuss, sharing his knowledge and expertise with me and contributing to improve my work. Thanks also for hosting me during my stay at the Virtual Acoustics Group at Aalto University.

I want to thank Prof. Jens Blauert for reviewing the work and for his expert advice that helped improve the quality of this thesis.

During these years I had the opportunity to collaborate closely with Dr. Winfried Lachenmayr both in Detmold and during my stay at Müller-BBM. I want to thank him and Dr. Eckard Mommertz for hosting me during my visit and for the nice work atmosphere that I enjoyed there. I'm thankful to the office colleagues from Aalto University, who welcomed me during my time there, and to Dr. Jukka Pätynen for nice collaborative work.

I am thankful also to colleagues from HfM Detmold. Thanks to Dr. Timo Grothe for the assistance and scientific discussions along the way, and for providing trumpet directivity data. Thanks to Malte Heins, for being always ready to assist with technical problems, to Gaby Aßmann for administrative support and to all the students and professors of the Erich-Thienhaus-Institut, which has been my second home during these years. Acoustic students have been great colleagues during this time, but they have become also good friends. Thanks to musicians and tonmeisters for being there during long recording and listening sessions.

Fellows of the BATWOMAN consortium have been a great support during these years, discussing and learning from them (each time in a different part of the world) has been a very enriching experience. Thanks also to my friend M.Sc. Francesc Capó for fruitful discussions on the analysis of datasets.

Finally I want to give my most sincere and deep thanks to my family. In special, I want to thank to my parents, Sebastià and Antònia, for supporting me during all my life and encouraging me to achieve all my goals. Last but not least, thanks to Paola, for joining me in this journey and making it also hers, for all the scientific discussions that we had during these years, and for her infinite patience.

Glossary

Abbreviations

3D	Three Dimensional
AD	Analog-to-Digital
ANOVA	ANalysis Of VAriance
ASR	Average-to-Silence Ratio
ASW	Apparent Source Width
BPM	Beats Per Minute
BRIR	Binaural Room Impulse Response
BTL Model	Bradley-Terry-Luce Model
BS	Brahmssaal
C_{50}, C_{80}	Clarity
CAD	Computer-aided Design
CTC	Cross-Talk Cancellation
D3S	Detmold Surround Sound Sphere
DA	Digital-to-Analog
DirAC	Directional Audio Coding
dB	Decibel
DSP	Digital Signal Processing
DST	Detmold Sommertheater
DTW	Dynamic Time Warping
ED100	Early-to-Direct Energy Ratio
EDC	Energy Decay Curve
EDT	Early Decay Time
EEL	Early Ensemble Level
ER	Early Reflections
FDN	Feedback Delay Network
FDTD	Finite-Difference Time-Domain
FEM	Finite Element Method
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FOA	First Order Ambisonics
FRF	Frequency Response Function
G	Sound strength
GA	Geometrical Acoustics
GPU	Graphics Processing Unit

GUI	Graphic User Interface
HRTF	Head Related Transfer Function
Hz	Hertz
ILD	Inter-aural Level Difference
I/O	Input-Output
IR	Impulse Response
ISO	International Organization for Standarization
ITD	Inter-aural Time Difference
ITU	International Telecommunication Union
J_{LF}	Early Lateral Energy Fraction
JND	Just Noticeable Difference
KH	Detmold Konzerthaus
L_J	Late Lateral Energy
LRA	Loudness Range
LTI	Linear Time Invariant
LUFS	Loudness Units relative to Full Scale
MFA	Multiple Factor Analysis
MFCC	Mel Frequency Cepstral Coefficient
MIDI	Musical Instrument Digital Interface
MIR	Music Information Retrieval
MIREX	Music Information Retrieval Evaluation eXchange
MPA	Music Performance Analysis
MRIR	Multi-channel Room Impulse Response
NR	Noise Ratio
RES	Reverberation Enhancement System
RIR	Room Impulse Response
RMS	Root-Mean-Square
RT	Reverberation Time
SDM	Spatial Decomposition Method
SF	Spectral Flux
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRIR	Spatial Room Impulse Response
ST_{early}	Early Stage Support
ST_{late}	Late Stage Support
ST_{total}	Total Stage Support
TDOA	Time Difference of Arrival
T_s	Center Time
VAE	Virtual Acoustic Environment
VBAP	Vector Base Amplitude Panning
VPS	Virtual Performance Studio
VSS	Virtual Singing Studio
WFA	Wave Field Analysis
WFS	Wave Field Synthesis
ZC	Zero Crossing

Musical concepts

Dynamics	Musical term used to refer to the playing level.
Dynamic Variations	Variations on the playing level over time.
Piano	Term referring to soft dynamics.
Forte	Term referring to loud dynamics.
Tempo	Term referring to the speed of playing of a piece.
Tempo variations	Variations on the speed of plying over time.
Articulation	Term used to refer to define the technique affecting the transition between consecutive notes.
Staccato	Term referring to a detached articulation or short playing of consecutive notes.
Legato	Term referring to a smooth articulation between consecutive notes.
Vibrato	Expressive technique consisting of a regular change of pitch, applied on a sustained note.

Contents

1 Introduction

From a pure technical perspective, music is the combination of sound and silence, the presence and the absence of sound pressure changes. However, in practice, the combination of distinct rhythms, timbres, pitches and amplitudes results into something much more complex, capable of transmitting expression, mood and emotion [Jus13b]. Music is one of the main sources of enjoyment and entertainment, and although nowadays a big portion of music consumption is done through recordings, live performance can be considered the greatest representation of music.

Musician, instrument, room and listener are the basic components of a western classical music performance. The actions executed by a musician on their instrument lead to the generation of sound, which is transformed by the room before reaching the ears of a listener. However, during a musical performance, musicians are as well active listeners, that interact with the sound that generated through their instruments and the alterations caused by room acoustics.

To accommodate room acoustics and suit their musical interpretation to the present space musicians can adjust their performance and playing style. This process results into the creation of a circulative feedback loop, where the sound generation is constantly shaped and modified by the actions of the musicians according to their perception. For instance, a musician could adjust several interpretative aspects e.g. dynamics, articulation, *tempo*... - in order to fulfill the aesthetic and musical demands a piece by taking advantage of the present acoustic conditions. Contrarily, room acoustics could affect as well negatively to the interpretation of a musical piece, forcing the musician to mitigate the effects of adverse acoustics by modifying the playing style.

The main focus of this work is to characterize potential performance adjustments made by musicians under different room acoustic conditions, and to evaluate the perceptual impact of these potential adjustments from a listener perspective. To this end, a virtual acoustic environment is implemented, allowing the conduction of formal studies with semi-professional musicians in controlled acoustic conditions. These studies include stage acoustic preference tests, recordings and analysis of live performances under different acoustic conditions, and perceptual evaluation of performance adjustments.

1.1 Scope

In order to approach the research question intuitively, a schematic view of the problem is presented in Fig. 1.1. A main circulative feedback loop is composed by the musician, their instrument, and the room. The mechanical actions executed by the musician on their

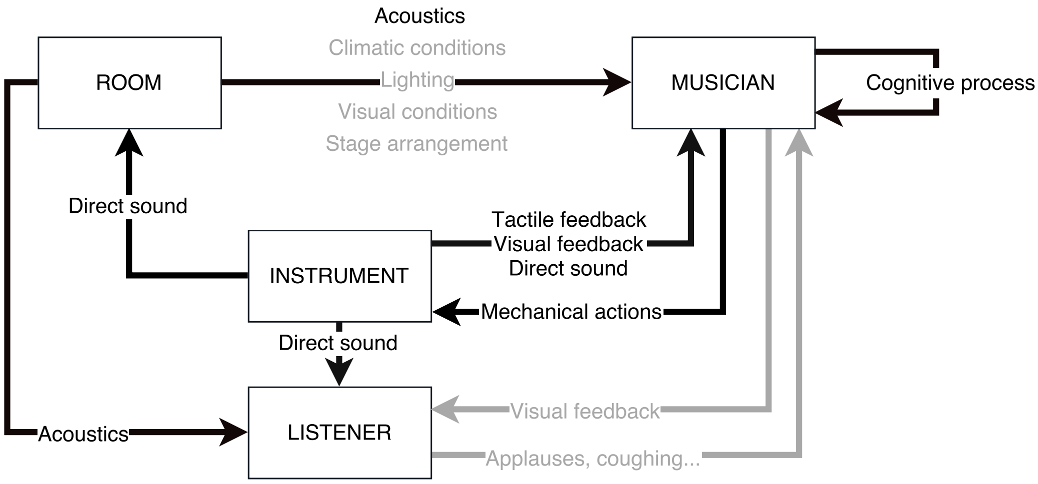


Fig. 1.1.: Schematic representation of the research problem.

instrument lead to the generation of sound. This sound is then modified by the acoustic conditions of the room and received again by the musician, who potentially reacts to the received sound by adjusting their performance. Other environmental conditions of the room, such as climate, lighting, visual aspects, or stage arrangements, are as well potential sources of influence on the performance. The reaction of a musician is governed by a cognitive process, which evaluates the acoustic and environmental conditions. Such process is affected by many factors, such as psychological and physical state, or preconceived concept of performance, among others. Alien to this situation, a listener acts as a passive element in this process, although limited feedback from listeners e.g. applauses, coughing – can be transmitted the musician. Parts of the problem that are addressed in this work are represented in black color, while parts represented in grey are out of the scope of this research.

Due to the interdisciplinary nature of the problem, the research has been divided into four basic modules, as organized in Fig. 1.2. The first module deals with the creation of a virtual environment capable of reproducing the acoustic feedback of performance spaces in real-time and in controlled conditions. This allows a musician to perform under different acoustic conditions by means of an electro-acoustic system. Once a virtual environment is implemented, formal studies can be systematically completed using controlled acoustics. The second research module focuses on studying musicians' preferences to room acoustics, putting special focus on their perception of acoustics under different performance contexts. This can provide important insight regarding the cognitive process that governs musicians' reaction to room acoustics. Then, the result of musicians' perception to acoustics and their reaction potentially translates into an adjustment of their performance, which is studied in the third module. These adjustments are investigated by conducting recordings of musical excerpts where musicians perform under distinct acoustic conditions. The final step is to evaluate the perceptual aspects of the mentioned potential performance adjustments, and analyze the perceptual relevance from the listener perspective.

The described problem leads to the formulation of a series of research questions focused on live performance of solo musicians:

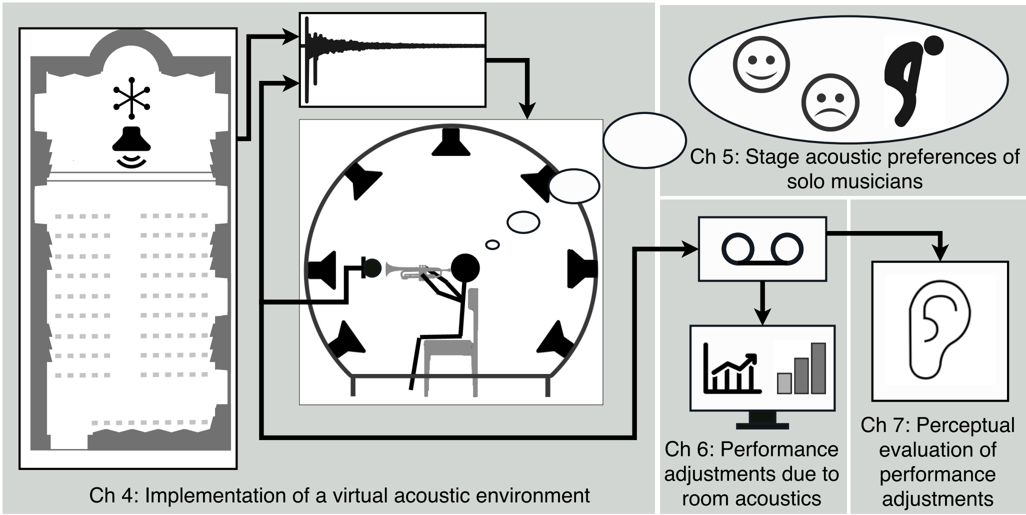


Fig. 1.2.: Schematic representation of the project structure.

- Is the usage of virtual acoustic spaces to systematically study live music performance appropriate? What are the advantages provided by this approach, as compared to traditional *in-situ* studies?
- How do room acoustic conditions affect to musicians in terms of taste, comfort and psychophysical state in different performance contexts?
- Do musicians adjust their performance depending on the acoustical conditions of a room? If so, can these adjustments be characterized and classified according to individual subjects, instruments and musical pieces?
- Are the potential adjustments performed by musicians perceivable by a listener?

1.2 Structure of the thesis

Chapter 2 includes the background information related to measurement and quantification of room acoustics, as well as existing systems to auralize room acoustic conditions in real-time. The chapter reviews the state of the art in research fields related to the study of room acoustic effects on musical performance and musicians' stage acoustic preferences.

Chapter 3 reviews the methods used to conduct systematic research studies in virtual environments with musicians. It presents the techniques used to implement virtual acoustic environments, as well as the methodologies used to implement automatic analysis of musical performance using MIDI and audio data.

Chapter 4 presents the implementation of a real-time auralization system. The Detmold Surround Sound Sphere (D3S) is a listening environment implemented during this project to allow the real-time auralization of room acoustics for musicians. The auralized rooms are measured using a microphone array and resynthesized using spatial analysis and reproduction techniques. The environment allows a musician to perform inside a 3D loudspeaker set-up while experimenting the resynthesized acoustic conditions of measured rooms.

Chapter 5 investigates stage acoustics preferences of solo trumpet players. Pilot studies with trumpet players performing in virtual acoustics resynthesized using the D3S system are introduced. Two studies are presented: a first one explores the effect of performance context on the preferences of trumpet players, meaning that the same room conditions are evaluated under different case scenarios – practice of instrument technique, practice of a concert piece, performance of a concert, ease of playing, and overall acoustic quality. The second study evaluates the impact of the direction of early energy on the acoustic stage support perceived by musicians. The results of the studies, together with feedback from the musicians are used to establish relationships between their physical and mental state and room acoustic conditions.

Chapter 6 investigates the performance adjustments implemented by musicians when playing under different acoustic conditions. Again, two studies are presented, one centered on the performance of trumpet players completed using the D3S system, and a second one where organ players are presented with enhanced acoustic conditions in the Detmold Konzerthaus, which is equipped with an electro-acoustic system that allows the modification of the reverberation time. The procedure followed in both studies consists of recording sessions with musicians where they are asked to prepare musical excerpts that are later recorded under different room acoustic conditions. The recordings of trumpet players consist of audio files, while the organ players are recorded using a MIDI interface installed on the organ. The recordings are then evaluated using automatic extraction of musical features, and the performance adjustments are characterized.

Chapter 7 contains the description of listening tests which evaluate organ and trumpet recordings obtained during the experiments from the previous chapter. Listeners are asked to judge and rate the differences between recordings obtained under different acoustic conditions.

Chapter 8 presents a discussion of the findings and final conclusions, as well as proposals for future work.

2 Background

This chapter presents a review of the literature related to the main topics of this thesis. Due to the interdisciplinary nature of the different parts of the work, the chapter has been split into three main sections. The first part covers the literature related to room acoustics, presenting the main room acoustical parameters and a summary of findings related to musicians' preferences on stage acoustics. The second section covers a review of existing spatial audio techniques for measurement and reproduction of room acoustics and the main characteristics of existing real-time virtual acoustic environments. The last part is related to the musical performance, including a review of the most important musical aspects, common ways to automatically analyze solo performances and a summary of results from previous studies on the effect of room acoustics on live music performance.

2.1 Room acoustics

2.1.1 Room Impulse Response (RIR)

A Linear Time Invariant system (LTI) can be fully described by an Impulse Response (IR), which describes the behavior of a system when it is fed with a Dirac delta. Applied to acoustics, a room can be understood as a set of LTI systems (one for each source-receiver position). The point-to-point acoustic response of a room can be fully described using a Room Impulse Response (RIR). A RIR is composed by the direct sound, a set of early reflections and a late reverberation tail (see Fig. 2.1).

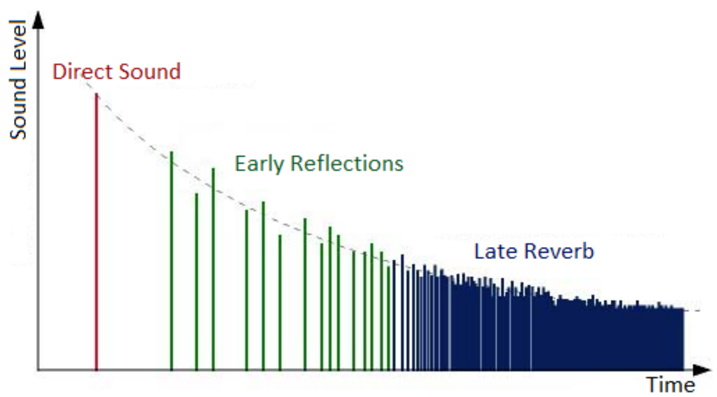


Fig. 2.1.: Example of a Room Impulse Response.

Additionally, an ideal RIR can be understood as a sum of individual acoustic events, which can be direct sound, specular and diffuse reflections, and diffractions:

$$p(t) = \sum_{n=0}^{N} p_n(t) \tag{2.1}$$

where $p(t)$ is a pressure RIR, t is time and $p_n(t)$ is the sound pressure of every individual acoustic event. Note that in a measured RIR there is always presence of noise. In addition, if every acoustic event is associated with a direction of incidence, a RIR can be parametrized, generating a Spatial Room Impulse Response (SRIR), which describes the spatio-temporal acoustic behavior of the room:

$$h_{\theta,\phi}(t) = \left[\sum_{p=0}^{P} h_{(p,\theta,\phi)}(t) \right] \tag{2.2}$$

where θ corresponds to azimuth and ϕ is elevation in a spherical coordinates system.

The transition time between early reflections and late reverberation - where the energy is dominated by a diffuse soundfield, is known as mixing time. The mixing time depends on the geometrical and absorption characteristics of the room, and it is a function of the echo density, D_e [DP08]. For simplicity, in room acoustic measurements 80 ms is normally considered the end of the early reflections and the start of the late reverberation.

$$D_e(t) = 4\pi c_0^3 \frac{t^2}{V} \tag{2.3}$$

where D_e is the echo density, c is the speed of sound, V is the volume of the room, and t is time.

2.1.2 Room acoustic parameters

Room acoustic parameters are used to describe quantitatively the acoustic properties of a space. These parameters are often computed using monaural RIR, providing information about the temporal and spectral behavior of the sound energy. In addition, a binaural RIR (BRIR) allows the computation of binaural parameters and a multichannel RIRs opens the possibility of computing parameters related to the spatial properties of the sound. The most common room acoustic parameters are listed in the standard ISO 3382-1:2009 [ISO09], which details the technical requirements of the acoustic measurements, as well as the procedure to derive the room acoustic parameters from RIRs. A list of perceptual attributes which relate to the standard parameters is provided as well. However, most of the parameters described in the standard are monaural parameters, with only a few of them accounting for the spatial properties of the sound.

Reverberation Time (RT$_{60}$) is the most common room acoustic parameter. It represents the time needed for the sound energy to decay 60 dB in an enclosed space after a sound source has been stopped. An estimation of the RT_{60} can be computed using the Sabine/Eyring equation, provided that the volume of the room and the absorption properties and surface of the materials are known.

$$RT_{60(Sabine)} = \frac{0.161V}{\sum_{i=1}^{n} S_i \alpha_i + 4mV} \tag{2.4}$$

where V refers to the total volume of the room (in m^3), S_i and α_i refer respectively to the area (in m^3) and absorption coefficient of the surfaces of the room, and m is the air absorption. Given that the necessary quantities to compute the RT_{60} are often unknown, a reliable method to estimate the reverberation time is based on measuring the decay time of 60 dB on the Energy Decay Curve (EDC) of a measured RIR. However, if the decay of the EDC is smaller than 60 dB, the reverberation time can be estimated from the extrapolation of the decay time over smaller ranges using a linear fit. The most common estimations are T_{20} and T_{30}, where the level ranges are from -5 dB to -25 dB and -35 dB, respectively.

Early decay time (EDT) is defined as the best-fit linear regression of the first 10 dB of decay of the EDC. It is related with the perceived reverberance of a room.

Clarity (C_{50}, C_{80}) defines the balance between early (up to 50 or 80 ms) and late arriving energy. The parameter relates to the perceived clarity of sound.

$$C_{50} = 10 \cdot \log_{10} \left[\frac{\int_0^{50} p^2(t)dt}{\int_{50}^{\infty} p^2(t)dt} \right] (dB) \tag{2.5}$$

$$C_{80} = 10 \cdot \log_{10} \left[\frac{\int_0^{80} p^2(t)dt}{\int_{80}^{\infty} p^2(t)dt} \right] (dB) \tag{2.6}$$

where p is the instantaneous pressure of the RIR at the measurement position. The time integration limits of the expressions are expressed in ms.

Sound Strength (G) is used to parametrize the amount of amplification provided by a room with respect to a source in free field. It is related with the subjective level of the sound in a room.

$$G = 10 \cdot \log_{10} \left[\frac{\int_0^{\inf} p^2(t)dt}{\int p_{10}^2(t)dt} \right] (dB) \tag{2.7}$$

where p_{10} is the the instantaneous sound pressure of the impulse response measured at a distance of 10 m in a free field, using the same sound source.

Centre Time (T_S) is the temporal center of gravity of the squared impulse response. T_S is related with the perception of clarity.

$$T_S = 10 \cdot \log_{10} \left[\frac{\int_0^{\infty} tp^2(t)dt}{\int_0^{\infty} p^2(t)dt} \right] (s) \tag{2.8}$$

Early Lateral Energy fraction (J_{LF}) measures the strength of the early lateral energy in a room. It relates with the apparent source width (ASW) perceived by the listener in the audience area. The measurement procedure requires the use of an omnidirectional and a figure of eight microphone at the same position. The parameter is computed by dividing the early lateral energy (from 5 to 80 ms, captured by the figure of eight microphone) by the total energy contained in the first 80 milliseconds:

$$J_{LF} = \frac{\int_{5}^{80} p_L^2(t) cos^2\theta dt}{\int_{0}^{80} p^2(t) dt} \tag{2.9}$$

Late Lateral Energy (L$_J$) follows a similar approach as J$_{LF}$, but in this case the integration times are different, taking 80 ms as starting time, thus computing the lateral strength of the late reverberation. The reference signal in this case is the pressure in free field at 10 m. L$_J$ is related to the subjective perception of envelopment.

$$L_J = 10 \cdot \log_{10} \left[\frac{\int_{80}^{\infty} p_L^2(t) cos^2\theta dt}{\int_{0}^{\infty} p_{10}^2(t) dt} \right] dB \tag{2.10}$$

Besides the previously presented parameters, there are a number of stage acoustic parameters which aim at the analysis of stage acoustics from the perspective of musicians. The most used parameters are early and late stage support, which were defined by Gade in [Gad89b].

Early Stage Support (ST$_{Early}$) is used to asses the ensemble conditions on stage.

$$ST_{Early} = 10 \cdot \log_{10} \left[\frac{\int_{20}^{100} p^2(t) dt}{\int_{0}^{10} p^2(t) dt} \right] dB \tag{2.11}$$

Late Stage Support (ST$_{Late}$) relates to the perception of reverberance on stage.

$$ST_{Late} = 10 \cdot \log_{10} \left[\frac{\int_{100}^{1000} p^2(t) dt}{\int_{0}^{10} p^2(t) dt} \right] dB \tag{2.12}$$

Total Stage Support (ST$_{total}$) is related to the support of the room to the sound of a musician's own instrument.

$$ST_{Total} = 10 \cdot \log_{10} \left[\frac{\int_{20}^{1000} p^2(t) dt}{\int_{0}^{10} p^2(t) dt} \right] dB \tag{2.13}$$

According to the standard [ISO09], in all cases the ST parameters are to be measured with the microphone positioned at a distance of 1 meter from the acoustic source.

Early Ensemble Level (EEL) was proposed by Gade [Gad89b] and it is defined as the ratio between the early energy and the direct sound at a distance of one meter. However, the parameter was later omitted [Gad92], since it presented no correlation between its values and perceptual attributes.

$$EEL = 10 \cdot \log_{10} \left[\frac{\int_{0}^{80} p^2(t) dt}{p_{dir}^2(t) dt} \right] dB \tag{2.14}$$

where p_{dir} refers to the pressure of the direct sound.

Modifications of stage parameters – G and ST – were suggested by Kato *et al.* [Kat+15]. The proposed parameters are G$_{early}$, G$_{late}$, and G$_{total}$ - relative early, late and total sound

	Subjective aspect	Freq. avg. (Hz)	JND	Typical range
Audience parameters				
Early Decay Time (EDT)	Reverberance	500 to 1000	Rel. 5%	1 s to 3 s
Clarity (C_{80})	Perceived clarity	500 to 1000	1 dB	-2 dB to 10 dB
Strength (G)	Level of sound	500 to 1000	1 dB	-2 dB to 10 dB
Center Time (T_S)	Perceived clarity	500 to 1000	10 ms	60 ms to 260 ms
Early Lat. Energy Frac. (J_{LF})	Apparent source width	125 to 1000	0.05	0.05 to 0.35
Late Lat. Sound Level (L_J)	Listener envelopment	125 to 1000	Unknown	-14 dB to 1 dB
Stage parameters				
Early Stage Support (ST_{Early})	Ensemble conditions	250 to 2000	Unknown	-24 dB to -8 dB
Late Stage Support (ST_{Late})	Reverberance	250 to 2000	Unknown	-24 dB to -10 dB

Tab. 2.1.: Standard room acoustic parameters according to ISO 3382 [ISO09].

strength, respectively. The main advantage of them is that they do not require the use of a free field measurement at 10 m, decreasing significantly the complexity of the measurement.

$$G_{early} = 10 \cdot \log_{10} \left[\frac{\int_0^{100} p^2(t)dt}{\int_0^{10} p^2(t)dt} \right] \tag{2.15}$$

$$G_{late} = 10 \cdot \log_{10} \left[\frac{\int_0^{10} p^2(t)dt + \int_{1000}^{\infty} p^2(t)dt}{\int_0^{10} p^2(t)dt} \right] \tag{2.16}$$

$$G_{total} = 10 \cdot \log_{10} \left[\frac{\int_0^{\infty} p^2(t)dt}{\int_0^{10} p^2(t)dt} \right] \tag{2.17}$$

All the presented parameters are usually computed in octave bands. To this end, the RIR is decomposed into frequency bands using a filter bank, and applying the equations on every band-limited signal. In order to reduce the amount of data, the standard ISO 3382 [ISO09] establishes guidelines for the averaging of the parameters over the different frequency bands, as well as the just noticeable differences (JND) and typical ranges of the parameter values. A summary of the standard parameters is presented in Tab. 2.1.

Although the measurement procedures and standard parameters of the ISO 3382 provide a straightforward manner to compare results among different halls, in the last years the need for a revision of the parameters and measurement procedures has been growing among the scientific community [Lok13; AP12; Kah16; BR15]. The main claims are that standard room parameters represent an overly simplified characterization of the acoustic properties of the room. The parameters are measured on several positions and computed in several frequency bands, but are afterwards spatially and frequency averaged. In addition, omnidirectional sources do not represent the radiation characteristics of any of the existing instruments, yet it is a requirement to use them when performing standardized measurements. Directional sources, such as the loudspeaker orchestra proposed by Pätynen *et al.* [Pät11], represent alternative approaches able to replicate spectral and directional characteristics of a full symphonic orchestra. Finally, standard parameters do not account for spatial characteristics of the sound-field, while the human hearing system presents several spatial dependent non-linear behaviors [Bla96].

There have been some attempts to formulate spatial stage parameters such as Directional Stage Support [Cab+10] or top-side spatial ratio for the early energy [Pan+17]. Additionally, standard parameters can be presented as a function of the angle (spatial dependent parameters) [Gut+13]. However, the measurement techniques used to derive those parameters are not standardized, difficulting their implementation and calibration.

2.1.3 Stage acoustic conditions and musicians

The investigation of concert hall acoustics has been a topic under research for the last century, and has been mostly focused on the acoustic conditions for the audience. However, stage acoustics contribute to shaping the sound of musicians through absorption, specular reflections, diffusion and diffraction, thus affecting how musicians perceive the sound of their own instrument as well as the sound of fellow players. In addition, stage acoustics are part of musicians' working conditions, hence impacting the health and in many cases inducing hearing losses on them. Although usually the hearing thresholds of classical musicians are not close to clinically significant levels [Käh+04], violin and viola players show higher hearing thresholds on the left ear, as a result of the instrument position and the difference in sound between their ears [Roy+91]. In addition, brass and percussion players of a ballet orchestra tended to show higher hearing threshold than other orchestra players, due to the high level of their instruments [Rus+13]. Besides health related issues, acoustic conditions on stage affect to comfort of musicians and communication across the stage, thus influencing the musical performance [Mey95]. For these reasons, research on stage acoustic conditions has become more common since almost four decades ago. This section focuses on findings related with solo musicians and conditions for small ensembles.

In 1978, Marshall *et al.* [Mar+78] conducted a study investigating the preferred acoustical conditions of string trios. The musicians were recorded in anechoic conditions and then they were asked to play individually along with recordings of the other players. Artificial single reflections with different timing, level and direction were added to the played recordings, and musicians had to state their preference among the presented scenarios. Although the acoustic situations were highly simplified, Marshall was able to state that there is a time window in which reflections from other players should arrive, specifically between 17 and 35 ms. From the results of this experiment, Marshall implemented a second test in which multiple reflections laying in this time window were played together with the direct sound of the other two members of the string trio. Those reflections were then high or low filtered, in order to establish whether a timbral difference affects to the ease of ensemble. It was found that high-passed reflections in general contributed more to ease of ensemble than low passed reflections, and results suggested that the ensemble bandwidth is focused in the range of 500 Hz to 2 kHz. Musicians could perceive timbral changes depending on the filtering, and in most cases had an aesthetic impact that could indeed influence their judgment. Similar studies on the delay of a single reflection were conducted later by Nakayama [Nak84]. Alto recorder players were asked to rate the ease of playing. He concluded that the most preferred delay time of a single reflection depends on the auto-correlation function (ACF) of the played music motif and the amplitude of the reflection. Comparable experiments were repeated by Sato textitet al. in 2000 [Sat+00] with cello players, reaching the same conclusion as Nakayama. Although those findings provide a preliminary knowledge for the design of stage enclosures, the sound-field on stage is much more complex, and the acoustic preferences of musicians (and listeners) are multi-dimensional [KL17; Kah16].

Another series of early experiments were conducted by Gade in 1989 [Gad89b] with solo players and duos in laboratory conditions. By mixing the sound of the real instruments with delayed versions and the sound reproduced in an anechoic room, different acoustic conditions were generated. The set-up was implemented with solo musicians and duos. The conclusion of the study was that soloists prefer audible levels of early reflections, which could be described as *support*. However, for certain instruments the early energy could be masked by the direct sound of the instrument. In addition, a higher level of late reverberation was preferred by the musicians. Regarding ensembles, he stated that the delay of the direct sound between players should be rather small, and the level of the early energy contributes positively to hearing each other. Contrarily, late reverberation could contribute negatively. These experiments led as well to the proposal of two of the previously presented stage acoustic parameters: Stage Support (ST), and Early Ensemble Level (EEL). In a study conducted in Danish and British concert halls [Gad89a] Gade analyzed the correlation between room and stage parameters and perceptual aspects related with ensemble performance. The Danish halls presented significant correlations between perceptual aspects and stage parameters (ST_{Early}, ST_{Late}), while no significant correlations were found for the British halls. Instead, the strongest correlations found for British halls were related to reverberation time (RT_{60}, EDT) and measures of clarity (T_{S}, C_{80}). Gade associated these differences to the individuality of performers and the familiarity of the musicians with the rooms, who could tend to be more affected by the perceived reverberance in case of not being used to perform in a specific hall. Another outcome of those studies was a first description of the multidimensionality of the stage acoustic requirements. In those halls, two perceptual dimensions accounted for more than 90% of the variance in responses. A first dimension dominated by ensemble attributes, support and reverberance accounted for more than 80% of the variance, while approximately a 10% corresponded to timbre.

In [UT03], Ueno *et al.*, conducted preference experiments using resynthesized soundfields with 13 professional music players (4 violins, 3 violas, 3 flutes, 2 oboes and 1 clarinet). The experiments consisted on rating the preference of different modified sound-fields where three parameters were systematically modified: levels of early reflections, length of the reverberation and the level of one late reflection. The results suggested that all three parameters affect to the subjective impression of stage acoustics, and acoustic support for soloists is provided by late reflections rather than early energy. Among the presented scenarios with three different reverberation times, it was found that wind players preferred reverberation times around 1.9 s, while string players preferred longer reverberation (2.2 s). Shorter times did not help to hear the room, while longer times could affected negatively to hearing other players. A following study [Uen+05] was conducted implementing the same sound-field modifications for the study of duos, with the participation of 14 professional musicians. The study concluded that early energy contributes to hearing the sound of the co-players, although an excessive level can be harmful. In addition, it was found that the level of the reverberation was more important than its length, and it was related with the ability of hearing the other musician and the ease of "making harmony".

Chiang *et al.* [Chi+03] conducted in-situ experiments in 5 halls with four different stage conditions in each hall (two seating arrangements and two side reflectors configurations). Three solo players (piano, violin and trumpet), a duo (violin and piano), a trio (violin, cello, and piano) and a brass quintet evaluated and commented on the stage conditions. It was found that brass players and the piano player preferred strong late energy, and the within-hall and among hall variations were comparable. In addition, while the frontal position without side reflectors was unsatisfactory for the cello player, it was preferred by trumpet players. Second order polynomial fittings were implemented relating the acoustical parameters and the subjectives preferences, finding that the most preferred value of early-to-direct energy ratio (ED100) is in the range of -12 dB to -11 dB, while most of the acoustic conditions with values in the range of -14 dB to -6 dB are generally acceptable for small ensembles and soloists. In addition, the most preferred value for ST_{late} is around -14 dB.

In [Jeo+14], Jeon *et al.* investigated solo, duo and ensemble preferences by in-situ comparison of several seating positions in two halls. The most preferred position to perform was located in the center of stage. It was found as well that stage parameters (ST_{early} and ST_{late}) and the temporal variation of the reflections correlated positively with the overall quality for solo musicians. Interestingly, it was found that ST_{Early} was negatively correlated with the ease of ensemble and the overall quality for quartets. However, only two halls were investigated, and as previously reported by Chiang *et al.* and Ueno [Uen+05] early energy contributes to hearing each other, but excessive energy can be harmful. Hence, it is possible that the values of the early energy in these two particular halls were overall too big.

While most instruments can be grouped into bigger categories (wind, strings, percussion), sharing common playing techniques and acoustic properties, the singing voice is the only one contained inside the human body, thus being unique. This affects indeed to the perception of the radiated sound, and as Fry stated in [Fry80], singers often need to rely on a second pair of ears, but when no external listener is available, singers' perception of their radiated sound depends strongly on the acoustic feedback provided by the room. In addition, after experiments with single simulated reflections, Noson *et al.* concluded that the lyrics of the songs have an influence on the preferred delays of early reflections [Nos+00; Nos+02]. Marshall and Meyer [MM85] conducted choir singing experiments in synthesized soundfields including the generation of early reflections and late reverberation. They concluded that the subjective aspects ease of singing and ease of ensemble were highly correlated. In addition, and in contrast with instrumentalists, late reverberation seems to contribute more to ease of ensemble for choir singers than early reflections. Nevertheless, early reflections within the first 40 ms are as well judged positively.

As a final remark, it should be noted that although the presented findings provide a solid basis to understanding the acoustic necessities of musicians on stage, there are a number of open questions at the moment. It seems that musicians' preferences present a certain level of individuality, which could be attributed to the sound production process, radiation characteristics and positioning of different instruments on stage. It is required thus, to perform systematic experiments focused on specific instruments or families in order to obtain more refined results. In addition, while many of the early studies constituted a starting point to investigate stage acoustic preferences, the plausibility and realism of the synthesized soundfields was often a matter of concern for the researchers [Gad89b]. With the current techniques it is possible then to implement research in topics with few previous work, such as the investigation of preference of directional sound-fields on stage.

2.2 Virtual acoustics

To investigate the acoustic stage preferences of musicians mainly two approaches are used: in-situ experiments in real performance halls, and laboratory experiments using sound-field resynthesis. Although initially it was a challenge to implement real-time resynthesized acoustic environments with acceptable plausibility and realism, nowadays it is possible thanks to the development of spatial analysis and reproduction techniques, in addition to the computational power of modern computers. Recently, it has become more common to implement a mixture of in-situ and laboratory environments by modifying the sound of performance rooms by electroacoustic means i.e. Reverberation Enhancement Systems (RES), and thus having the possibility to synthetically extend the response of a real room.

2.2.1 Auralization

Auralization can be described as the rendering of audible soundfields, as an analogy with visualization:

> "*Auralization is the process of rendering audible, by physical or mathematical modeling, the sound field of a source in a space, in such a way as to simulate the binaural listening experience at a given position in the modeled space."* - M. Kleiner in [Kle+93].

The aim of auralization is to recreate the acoustic impression of a space, either indoors or outdoors. Through this thesis, the term auralization will refer to the rendering of room responses. The basic scheme of auralization consists on capturing or generating a spatial room impulse response (SRIR) and convolving it with anechoic sound - live or recorded. Then, with appropriate spatial reproduction techniques, the result is played back to the listener, providing an immersive aural experience. The basic elements of an auralized scene are the sound source, the medium and the receiver. Depending on the target and the capabilities of a virtual acoustic environment (VAE), these elements can be static or dynamic. Dynamic environments usually require the implementation of real-time operations, while static environments can be stored as a multichannel audio file. Furthermore, environments allowing the interaction of musicians with the acoustic scene e.g. musicians performing in virtual rooms, require the real-time convolution of the live sound with a SRIR. Two main approaches can be used for the generation of SRIRs: acoustic measurements or computer simulations. The reproduction of SRIRs can be implemented using loudspeaker set-ups or headphones.

Obtaining SRIR from measurements requires the analysis of the soundfield at a specific location, in order to be able to parametrize and resynthesize it later. Common techniques to analyze soundfields using microphone arrays are Wave Field Analysis (WFA) [Ber+97], plane-wave decomposition [Raf03], Directional Audio Coding (DirAC) [Pul07], or Spatial Decomposition Method (SDM) [Ter+13]. In all these techniques, the main goal is to parametrize a soundfield, in order to be able to generate appropriate filters that can be used to convolve with anechoic sound. The parametrization of the soundfields strongly depends on the used reproduction technique.

Besides measurements, a soundfield can be computed using simulations, which can be grouped into two main categories: geometrical acoustics methods (GA) and wave based methods. The most common techniques in GA methods are Ray Tracing [OB89; Sav+99], Mirror-Image Source models [AB79; Bor84] or Beam Tracing, among others. It is frequent as well to combine various methods to implement hybrid models. An extensive review on geometrical acoustics modelling is provided by Savioja and Svensson in [SS15]. Wave based methods are focused on approximating the solution of the wave equation in a room. The most used techniques are Finite Difference Time Domain (FDTD) [Bot95] and Finite Elements Methods (FEM) [PK97]. At the moment, most of the wave based methods are limited to low and mid frequencies, and it is common to combine them with GA methods [MM13]. However, it is expected that in the following years simulations of the entire audible range will be performed in real-time using graphic processor units (GPU) [Sav10; WB11].

Spatial reproduction techniques are divided into two groups: loudspeaker based reproduction and headphone reproduction. The most common techniques based on loudspeaker reproduction are Vector Base Amplitude Panning (VBAP), Ambisonics, Wave Field Synthesis (WFS) and Cross-Talk Cancellation (CTC). Headphone reproduction is normally referred to as binaural reproduction.

Vector Base Amplitude Panning (VBAP) [Pul97] is a three-dimensional generalization of stereo. By using loudspeaker triplets with different amplitude, a phantom source is generated in the triangular plane conformed by three loudspeakers. The main advantage of VBAP is that can be implemented with arbitrary loudspeaker set-ups, although the sharpness of the localization depends on the distance between loudspeakers and the correct reproduction is limited to a sweet spot. Ambisonics reproduction is based on the description of a soundfield as a set of spherical harmonics (SH) [Ger73]. The combination of those SH results in the synthesis of a soundfield over a limited three-dimensional space. The order of the spherical harmonics is determined by the number of available loudspeakers for reproduction. Higher order results in a finer spatial resolution and normally it is necessary to use regular loudspeaker arrangements. Wave Field Synthesis (WFS) is based on the Huygens-Fresnel principle, and modeled using the Kirchoff-Helmholtz integral [Ber+93]. Although a perfect reconstruction can be theoretically achieved, the discrete nature of loudspeaker arrays leads to spatial aliasing above a certain frequency, which depends on the spacing between loudspeakers. In addition, the reproduction is limited to the horizontal plane. The main advantage of WFS is the suppression of the sweet spot, allowing the presence of multiple listeners in a larger area. Cross-Talk Cancellation (CTC) can be described as binaural reproduction using loudspeakers [Bau61]. In combination to the binaural signals, each loudspeaker plays a canceling signal to suppress the cross-talk between ears. While it provides sharp localization and in its dynamic implementation it can be independent from a sweet spot, the reproduction relies on an accurate tracking of the listener head position to deliver appropriate canceling signals [Len06].

Binaural recordings constitute a separate category, both in terms of capturing the soundfield and reproducing it. Binaural auralizations can be obtained in a very straightforward manner, if an artificial head is available. A simple static auralization consists of convolving anechoic sound with a Binaural Room Impulse Response (BRIR) or performing a binaural recording of live sound. The spatial impression of a binaural auralization relies on the correct reproduction of the two localization cues of the human auditory system: Interaural Level Difference (ILD) and Interaural Time Difference (ITD) [Bla96]. This implies that the correct spatial perception depends on the appropriate reproduction of these auditory cues. However, since every person has different head, ear pinna and ear canal shapes and sizes, binaural auralization based on artificial head recordings do not always provide satisfactory results. When this is the case, individualized measurements of head related transfer functions (HRTF) are necessary. The implementation of dynamic binaural scenes allowing head movements can be implemented with the combination of several BRIRs with different head orientation and the use of head-tracking devices. Finally, loudspeaker based reproduction methods can be virtualized over binaural reproduction by convolving the loudspeaker signals with HRTFs corresponding to the loudspeaker positions. This is useful to compare the spatial impression of different reproduction techniques without the necessity of several physical set-ups.

2.2.2 Virtual acoustic environments

The implementation of a virtual acoustic environment is normally a combination of different measurement or simulation and reproduction methods. This section gives an overview on environments previously implemented and used for the study of musical performance.

The early environments implemented by Marshall were based mostly on the reproduction of delayed repetitions of the direct sound generated by musicians in an anechoic room [Mar+78]. Later, in 1985, by Marshall and Meyer complemented the reproduction of early reflections with the inclusion of late reverberation [MM85]. That environment was composed of a digital delayline reproducing three first-order reflections corresponding to different sizes of a stage enclosure, complemented with two incoherent reverberation feeds generated by a reverberation plate (EMT Goldfoil). The reproduction set-up was composed of 7 loudspeakers, three of those reproduced early reflections (sides, rear and ceiling) and three reproducing reverberation (sides and front). A similar approach was taken by Gade in 1989 [Gad89b]. In this case only 5 loudspeakers were used (front, sides, rear, ceiling) to reproduce one early reflection and late reverberation (after 110 ms) generated by reproducing and capturing live sound of the musician in a reverberation chamber. The same procedure was followed as well to implement an acoustic environment for duos in two separate anechoic chambers. These examples represent early implementations of interactive auralization and virtual acoustic design, although the acoustic quality of those environments did not resemble the characteristics of real performance spaces and unnaturalness in timbre was often a matter of concern to the experimenters.

The availability of digital signal processing (DSP) tools supposed a great step towards the realization of plausible environments, specially with the implementation of real-time convolution. Ueno *et al.* [Uen+01] developed a measurement and reproduction method based on the multichannel RIR. Using a dodecahedral source on stage and a directional microphone six RIR are obtained at the same position by orienting the microphone towards orthogonal directions. The direct sound is then removed from the RIRs and the sound of a musician playing in an anechoic room is convolved with the measured RIRs. The reproduction set-up is composed of 6 loudspeakers (ceiling, floor, sides, front and back). Although the directional properties of the auralization were not evaluated the monaural parameters (RT, ST_{Early}, ST_{Late}) of the auralizations and the real rooms show a good agreement.

The Virtual Performance Space (VPS) [Lai+11] is an environment based on First-Order Ambisonics (FOA) reproduction of simulated RIRs using Ray-Tracing. A directional source and a receiver are arranged on stage of a simulated room, replicating the position of a musician and their instrument. The simulated soundfields are reproduced using a dodecahedral surround array of 12 loudspeakers. The naturalness of the auralization seems to be affected by factors such as movements of musicians, microphone positions, level of the reverberation and *PA Effect* (the feeling that the sound of the musician is amplified using a Public Address system).

The Virtual Singing Studio (VSS) [Bre+12; Bre14], is a system mostly addressed to the study of singing performances. The auralization procedure is based on the measurement of FOA room impulse responses using a studio monitor and a Soundfield microphone [Far79b; Far79a] in a real room. The reproduction is based as well on FOA using a set-up of 16 loudspeakers. The reverberation differences between the real and the auralized spaces are smaller than 10 % (twice the JND). The system was successfully used in experiments with singing voice performance.

Schärer implemented a virtual environment based on the auralization of Binaural Room Impulse Responses (BRIR) [SKW13a; SK15]. The BRIR were generated using computer simulations with directive sources and musician-instrument arrangements. Extra-aural headphones were used in the reproduction, in order to allow musicians to perceive the sound of their instrument. Using head-tracking it was possible for the musicians to move their head during the auralization. One of the key aspects for a binaural system is a correct headphone equalization [SKL09].

2.2.3 Reverberation enhancement systems

The common characteristic of most VAEs is that the acoustic properties of the target space are fully reproduced by electroacoustic means. However, it is possible as well to implement a Reverberation Enhancement System in a reverberant space, in order to enhance or modify the acoustic behavior of a room. The most common RAES are normally commercial systems which constitute an alternative to mechanical variable room acoustics.

Live sound in the room is picked up by microphones, digitally processed and finally played back to the room by a multichannel loudspeaker set-up. Although typically the implementation details of a commercial RAES are not available, one may intuitively think that the enhancement is based on Feedback-Delay Networks (FDN) [JC91], convolution of live sound with multichannel room impulse responses, or a combination of both methods. While RAES allow the user to modify room acoustics in a straightforward manner, the main limitation of the systems is that the decay of the room can only be extended.

Some of the available commercial RAES at the moment of writing this thesis are *E-coustic*, *Meyer Constellation*, *Müller-BBM Vivace*, *SIAP Acoustics*, *Wenger Virtual Acoustic Environment (VAE)*, and *Yamaha Active Field Control (AFC)*.

A *SIAP Acoustics* RAES is permanently installed in the Detmold Konzerthaus. In addition, a *Müller-BBM Vivace* system was temporarily installed and used during the work of this thesis.

2.3 Musical performance

Musical performance is a multidimensional and highly individual process, result of two main components, technique and expression [Slo00]. While the technical component is related to the motor skills involved in the execution of movements leading to the generation of sound, the expressive component is related to intentional variations in the performance with the finality of influencing the aesthetic and cognitive perception of listeners. C.E. Seashore and Metfessel stated that "the unlimited resources for vocal and instrumental art lie in artistic deviation from the pure, the exact, the perfect, the rigid, the even and the precise" (cited from H.G. Seashore [Sea37], p.155). The result of this is often the arousal of profound emotions [Jus13a]. In experimental research with performers and listeners, Gabrielsson [GJ96] found out that the expressive intention of a musician leads to variations on music characteristics e.g. tempo, dynamics, timbre - and the intended expression is generally perceived by a listener, although the process is influenced by individual strategies of the performers.

Various models of performers have been proposed in the literature, in an attempt to formulate the multiple variables that affect the musical performance. Schärer [SK15] proposes a performer model based on three main categories: the previous concept of a performance, the musical performance, and the potential influences from external variables. According to Gabrielsson [Gab99], the planning of a performance is a result of two interrelated processes: acquiring an adequate mental representation, and practicing to a satisfactory level. In addition, in Schärer's model, the musical notation on the score, personal style, performance tradition (or musical school) or personal style are other factors that influence the previous concept of a performance. This concept is translated into movements and actions during a performance, which result in the creation of sound. Auditory, tactile and visual feedback are cognitively processed by the musician, who reacts to this by continuously modifying their actions. External variables that could potentially affect to the musician are the personal form (physical and psychological state), the audience, and environmental variables. Room acoustics is an environmental condition that affects directly to the generated sound, and thus having a direct influence on the auditory perception of the musician.

Another performer model proposed by Ueno *et al.* [Uen+10] revolves around the same idea: the existence of a feedback loop between musician, instrument and room. Although the basic idea is similar, the model makes a distinction between common feedback and acquired feedback. The common feedback is presented as an automatic response, actions as a response to auditory sensations. On the contrary, the acquired feedback explains a more profound cognitive process in which a musician modifies their performance as a result of an imaginary image of the performance from the audience perspective, by combining a previous concept of the performance and the auditory perception. The behavior of this acquired feedback is indeed governed by the performing experience.

2.3.1 Music Performance Analysis (MPA)

Music Information Retrieval (MIR) [Dow03] is an interdisciplinary science with the objective of extracting relevant information from any kind of music related source, such as recordings, music scores or lyrics. Popular applications of MIR are identification of songs, recommendation, automatic music transcription, music genre classification or analysis and synthesis of musical expression, among others. In particular, the goal of Music Performance Analysis (MPA) is to study the characteristics of a musical performance by means of parameters that describe the generation and perception of music performance [Ler12].

A musical performance can be recorded in many ways e.g. audio, video, MIDI - allowing multiple approaches for the analysis and extraction of musical information. The analysis of audio data is commonly based on the extraction of audio features which describe certain low level characteristics of the recorded signal and the construction of perceptual models [Ler12; Fri+14]. Video recordings allow the analysis of gestures, movements and visual communication among players of an ensemble [Had12; Had+13]. Finally, a MIDI encoded stream contains relevant information regarding the interaction of a player with an instrument e.g. timing, pitch, level [Rot95].

As Seashore pointed [Sea38], there are four basic musical characteristics - time, pitch, loudness and timbre which relate to the physical properties of a wave: duration, frequency, amplitude and waveform. Friberg [Fri+11] proposes a limited list of relevant perceptual features for the communication of emotion during a musical performance which relate to the four characteristics proposed by Seashore. In this sense, time is explained by the tempo (slow-fast), rhythm (flowing-firm) and articulation (staccato-legato); pitch is explained by modality (major-minor) and overall pitch (low-high); loudness is explained by dynamics (soft-loud) and timbre is explained by brightness (dark-bright). Note that this selection does not intend to represent a complete list of perceptual musical features, but the most common ones.

One of the first devices to analyze music performance was the Iowa Piano Camera, by Tiffin and C. E. Seashore [TS30], developed in 1930. By using a camera pointing to the piano hammers, onset times, offset times and velocity of played notes were recorded on a photographic film. Another early invention by C. E. Seashore was the tonoscope [Sea14], a device used to determine the pitch of a sound. Based on the principle of stroboscopic vision, light projected on a rotating drum with a grid of small dots generated line patterns corresponding to the frequency of the sound.

The development of MIDI in the 1980s allowed a systematic approach to the extraction of performance information, specially with the commercialization of instruments with built-in MIDI recording devices, such as the Yamaha Disklavier or the Bösendorfer SE-System (later renamed CEUS) [Bol+94; Kaw+13]. The recording and reproduction accuracy of those instruments was studied by Goebl and Bresin in [GB03]. An organ with similar capabilities is installed in the Konzerthaus of the HfM Detmold. Nevertheless, the MIDI representations of musical performances are often limited to key instruments, and many instruments require the analysis of audio recordings to extract performance information.

Extracting onset and offset times from audio recordings is often more laborious. Alternatives to annotate timing information of a performance are tapping along while listening [DG02], manual annotation on audio edition programs or automatic extraction techniques. Automatizing the extraction of timing information can reduce significantly the effort of annotating time data, although it often needs to be complemented with a manual revision of the estimated times. The Music Information Retrieval Evaluation eXchange (MIREX) is an annual event consisting on a community based evaluation of MIR algorithms by comparing the results of different algorithms and systems applied to the same set of problems [Dow08]. For instance, the maximum average accuracy achieved in the MIREX 2016 blind onset detection challenge was 87% [DI16; SB14]. On the contrary, audio to score alignment using MIDI data as additional input data already shows very promising results, with a maximum average accuracy of 97% in the MIREX 2016 [DI16; CO+15]. The automatic extraction of pitch for monophonic signals presents often good results. A review of techniques is provided by Gerhard in [Ger03]. An appropriate estimation of pitch allows the study of vibrato and intonation in performances.

The extraction of low-level audio features - onsets, offsets, pitch, amplitude, spectral features, statistical parameters, etc. - is used to construct models that correlate physical signal features to perceptual aspects of performance. Using perceptual features as intermediate representations of a musical performance it is possible then to link low-level audio signal features or a combination of them to perceptual and emotional aspects of a performance [Fri+14]. Although the list of low level audio features is certainly large, summaries of description and implementation of relevant features for MPA can be found in [Ler08; Bre14; Pee04]. At the moment of writing this thesis there are as well a number of software solutions to extract performance parameters from both symbolic (MIDI) and audio recordings. Examples of these solutions are *jMir* [MF09], *MIR Toolbox* [LT07], *MIDI Toolbox* [ET04], *Essentia* [Bog+13], *Praat* [Boe01] or *Sonic Visualiser* [Can+10].

2.3.2 Influence of acoustics on live performance

As the previously described performer models state [SK15; Uen+10], the auditory sensation perceived by the musician is a variable that can potentially influence the performance. In this sense, the performer, their instrument and the room conform a feedback loop, meaning that the resulting performance is continuously adjusted. Although the modification of musical performance seems to be somewhat a generalized behavior among musicians, the adaptation strategies are highly individual. Variables that can affect the adaption are instrument, skill level of the performer or musical piece, although there could be more [SK15; Uen+10; Kat+15].

Piano is likely the most researched instrument in terms of performance, partly thanks to the analysis possibilities provided by built-in MIDI systems of certain piano models. During the

1990s, Bolzinger *et al.* completed various experiments with piano players using a Yamaha Disklavier. A preliminary experiment conducted in 1991 [BR92] investigated 5 players (4 professionals) in 4 acoustic situations with times ranging between 0.3 and 1.5 s. The variable acoustics were achieved using absorbing panels, and in order to isolate the variable under study, namely room acoustics, a paper curtain was present between the performer and the walls of the room, thus preventing the musician to perceive visual changes. It was found that aspects such as MIDI velocity (level of the pressed keys), duration of breaks, articulation and time of sustain pedal were varied as a function of the reverberation time. The general trend was that a higher reverberation resulted in softer, more dettached and slightly slower performances, with less use of the sustain pedal. In addition, it appeared that professional musicians varied their performance in a greater extent than the non-professional player. After that, Bolziger *et al.* completed a larger study recording 7 professional pianists in 8 different acoustic configurations [Bol+94]. In this case, and contrarily to the previous experiment, the tempo of the performance did not seem to be influenced by acoustic conditions. The overall intensity of the performance was found to correlate negatively to the reverberation time and level, meaning that pianists try to compensate the effect of the reverberation by playing softer. Although this study tried to isolate the effect of different room acoustic characteristics e.g. reverberation time, level and spectrum, and direct sound to reflections ratio (Dir/Ref) - it was not possible to ascribe specific effects to individual parameters.

A similar experiment was conducted by Kawai *et al.* with 12 professional piano players and MIDI recording analysis. The general trends among most musicians were in line with the results obtained by Bolzinger *et al.* finding strong negative correlations between ST_{Early} and note velocity mean and standard deviation. Additionally, the use of the sustain pedal was negatively correlated with the reverberation time. Finally, the overall tempo of the performance was not significantly affected. Besides the general trends, a preliminary cluster analysis showed that players could be classified into groups with similar behaviors. It is worth noting that in this case the experiment combined recordings in two virtual acoustic spaces and three real rooms, meaning that in some of the recordings, not only room acoustics but the entire environment was changed, which could indeed have an effect on the musicians' adaption strategies.

Ueno, Kato *et al.* [Uen+10; Kat+15] investigated the performance changes of 5 musicians (oboe, two flutes, baritone singer and violin) using the previously presented 6-channel virtual environment [Uen+01]. All the performers were asked to play the same two pieces with different tempos. It was found that for some performers very short and long reverberation times led to a decrease of the overall tempo. In addition, in some cases reverberation was positively correlated with the degree of *staccato*, and large reflection energy resulted in a decrease of level and harmonic strength of the produced sound. The *vibrato* characteristics of long notes was also analyzed, concluding that higher reverberation contributed to a more intense *vibrato* with lower frequency. Although the performance changes might be highly individual, the subjective reports on the adjustments often matched with the changes observed in the audio analysis. Moreover, some of the performance changes exceeded the JND of loudness, tempo, pitch and *vibrato* rate, leading to audible differences.

During her doctoral research, Schärer Kalkandjiev [SK15] completed a field study recording a professional cellist on tour [SKW13b] and a laboratory study with multiple musicians [SKW15]. The field study showed that more than 50% of the explained variance on the performance attributes were due to room acoustics. Very high and very low reverberation times lead to a decrease of the overall tempo. According to Schärer Kalkandjiev, while the cause of reducing the tempo in highly reverberant environments is most likely to reduce blurring between consecutive notes, the lack of reverberation in dry spaces forces the musician to prolong notes in order to replace the effect of reverberation, resulting on a decrease of the tempo. This changes were bigger on faster movements. Higher early and late energy resulted in a decrease of loudness, while longer reverberation contributed to an increase of loudness, meaning that the musician could perceive a difference between reverberant energy (ST_{late}) and duration of the reverberation (RT). Strong late energy was highly correlated with a brighter timbre with a larger bandwidth. Since it is likely that the instrument has an influence on the performance changes, in the laboratory study [SK15] Schärer conducted experiments with 12 professional players and 6 instruments (violin, cello, clarinet, bassoon, trumpet, trombone) using real-time binaural resynthesis of 14 simulated halls. In this case, the musicians were asked to play two pieces of their choice, one of them with fast tempo and the other one slow. It was then seen that the overall tempo changes were generally larger in the slow piece, reducing it in more reverberant environments. Contrarily to this, violin players tended to reduce the tempo of the fast piece and kept it constant for the slow one. Clarinet players did not show significant tempo differences. An increase of *Agogics* (tempo variations) was related mostly to early energy, although these changes seem to be highly piece and player dependent. While trombones and bassoons reduced the overall level of the performance in presence of higher early energy, other instruments did not react the same way, and trumpets even increased the level. Musicians often commented that good sounding rooms helped to increase the dynamics of the performance, and it was seen that cellos and clarinets tended to increase the dynamic range in environments with higher late energy (ST_{late}). However, other instruments did not follow the same trend. Finally, timbral changes of the generated sound were often related to spectral characteristics of the room (bass ratio, bass strength), suggesting that musicians could intend to compensate the tonal characteristics of the space.

A study with six amateur guitar players and singers [Hat12] reported strong correlations between the produced sound levels and standard room parameters. Strength (G) and reverberation time (RT) correlate negatively with the produced level, while higher clarity (C_{80}) results in an increase of the level.

As a summary of the previous findings it is clear that performers adjustments are highly piece dependent and instrument dependent. A generalization of the results is at the moment not available, since most of the studies dealt with small number of subjects and often the compared acoustical parameters are not the same in different studies. In spite of that, the presented independent studies on piano performance [BR92; Bol+94; Kaw+13] reported similar findings, indicating that a generalization for individual instruments could be achieved, at least to a given extent, and taking into account the individualities of every musician.

The fact that musicians playing the same instrument often exhibit different behaviors, suggests that the adjustments are as well highly individual and the physical and emotional state of the musicians, their musical background and experimental conditions could strongly influence the results. In addition, although most of the studies complemented the performance analysis with interviews, it is not clear to which extent the performance adjustments are conscious or intentional. Testing the same musicians in separated sessions could provide useful information regarding the influence of the personal state and environmental conditions on the results.

3 Methodology

The research methodology of this work is mainly based on the conduction of systematic experiments in controlled acoustic conditions, investigating and characterizing the effect of room acoustics on musicians. This chapter reviews the methods used to conduct the experimental research of the present work, and is divided into two main parts, auralization and performance analysis.

The techniques used to implement virtual environments based on the resynthesis of measured room impulse responses (RIR) are presented in Section 3.1. Section 3.2 presents the techniques used to analyze musical recordings, using both audio and MIDI formats. The features to characterize musical performances and the tools used for the analysis are presented and described.

3.1 Auralization

The auralization process used in this thesis is based on the measurement of SRIR in real rooms using a directional sound source and a microphone array. These SRIRs are then analyzed and resynthesized to generate appropriate filters. These filters are finally convolved in real-time with the live sound generated by a musician. A simplified block diagram featuring a generic auralization system is depicted in Fig. 3.1.

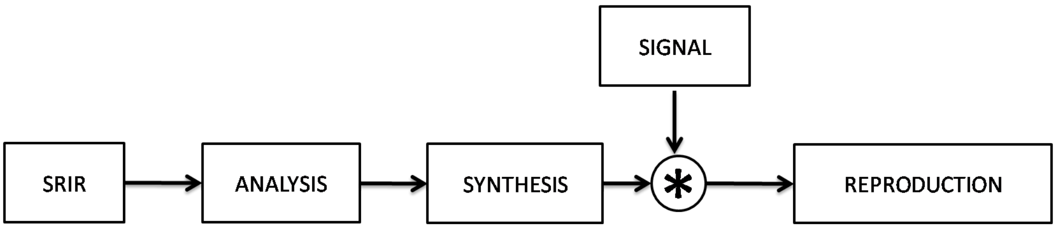


Fig. 3.1.: Main operations in a generic auralization process.

3.1.1 Acoustic measurements

The method employed for the measurement of RIR and multichannel RIR (MRIR) is the logarithmic sine sweep technique, developed by Farina [Far00]. This allows the simultaneous measurement of impulse response of a Linear Time Invariant (LTI) system and the distortion of a memoryless non-linear system. In this case, a room is a LTI system and a loudspeaker can present harmonic distortion and non-linear artifacts that are isolated from the linear response of the room.

The output signal of a LTI system is the convolution between the input signal $x(t)$ and the impulse response of the system $h(t)$ plus the presence of noise $n(t)$ which is assumed to be white, with Gaussian distribution and uncorrelated from the input signal.

$$y(t) = x(t) * h(t) + n(t) \tag{3.1}$$

Then, neglecting the presence of noise, usually much smaller than the level of the input signal, the impulse response of the system can be obtained by convolving the output by an inverse filter $f(t)$.

$$h(t) = y(t) * f(t) \tag{3.2}$$

The sine sweep deconvolution technique consists of using a logarithmic sweep with starting frequency w_1, ending frequency w_2 and duration T seconds as a system input, in this case, the signal played by the loudspeaker. In audio applications, the starting and ending frequency must cover at least the audible range, from 20 Hz to 20 kHz.

$$x(t) = \sin\left[\frac{\omega_1 T}{\ln\left(\frac{\omega_2}{\omega_1}\right)}\left(e^{\frac{t}{T}\ln\left(\frac{\omega_2}{\omega_1}\right)-1}\right)\right] \tag{3.3}$$

To generate the inverse filter $f(t)$, the excitation signal $x(t)$ is reversed in time and filtered using a filter with a decay of 6 dB/octave, having a level of 0 dB at ω_1 and $-6\log_2\left(\frac{\omega_2}{\omega_1}\right)$ at ω_2.

A deconvolved RIR presents a noise floor determined by the characteristics of the measurement equipment e.g. microphones, loudspeaker, amplifier. If the signal-to-noise ratio (SNR) of the measured RIR is not big enough, artifacts may be present when measured RIR are used for auralization purposes. The SNR can be increased by averaging several RIR measurements. However, in certain measurement situations it is not possible to sufficiently reduce the noise floor level. For this reason, the method proposed by Cabrera *et al.* in [Cab+11] is used in this work to suppress the noise floor in measured RIR, creating an infinite decay slope.

The denoise method basically consists of extrapolating the decay of the RIR in several frequency bands. First, the measured RIR is filtered into octave bands with center frequencies from 125 Hz to 16000 Hz. The resulting band limited RIR components are low pass filtered and transformed into logarithmic scale. The decay envelope of the RIR is then expressed as a combination of an exponential decay and a steady state noise floor and the following expression is fitted.

$$L(t) = 10 \log_{10} \left(10^{at/10} + b\right) \tag{3.4}$$

where $L(t)$ is the fitted level function in dB, t is time in seconds, a is the slope of the decay and b is the noise floor value (in linear scale). Then, a gain parameter $g(t)$ is applied to each band starting from the time t in which the level of the fitted envelope is 10 dB higher than the noise floor, meaning that the denoised impulse response is identical to the original up to that point.

$$g(t) = \sqrt{\frac{10^{at/10} + b}{10^{at/10}}} \tag{3.5}$$

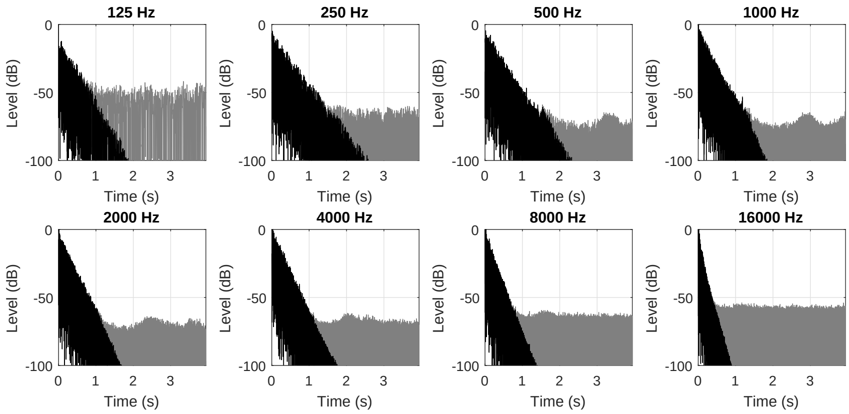


Fig. 3.2.: Original RIR (grey) and denoised version (black) filtered in octave bands.

Finally, a broadband RIR is generated again by summing all the denooised band limited RIRs. An example showing the original and denoised RIR is depicted in Fig. 3.2.

3.1.2 Spatial Decomposition Method (SDM)

The Spatial Decomposition Method (SDM) [Ter+13] aims at the extraction of directional information from MRIR in time domain. The main idea is that a pressure RIR can be parametrized by associating a direction of incidence to every sample of the response. The algorithm assumes that every sample represents only one single acoustic event, and thus a SRIR is represented as a set of consecutive plane waves or image sources.

An impulse response is captured with an array composed of at least 4 closely spaced microphones defining a three dimensional space. A small time window (typically in the order of 1 ms) is applied on the MRIR, and by using the least squares solution for time difference of arrival estimates (TDOA), the direction of incidence is estimated for a specific sample. Then, the window is shifted one sample and the directional analysis is repeated. This process is repeated for the entire MRIR. A block diagram showing the basic flow of the procedure is depicted in Fig. 3.3.

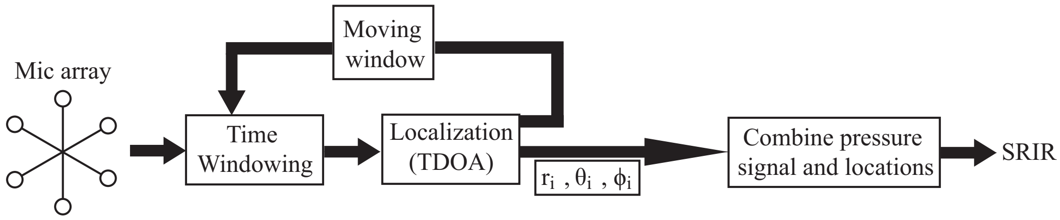


Fig. 3.3.: Block diagram of the analysis of a SRIR using the Spatial Decomposition Method.

The localization procedure is based on analyzing the time difference of arrival estimates (TDOA) of every sample at pairs of microphones. For an array of M microphones, the number of pairs is $\frac{M(M-1)}{2}$. The first step is to calculate the cross-correlation function of the windowed responses of every microphone pair.

$$
\begin{aligned}
\mathbf{R}_{1,2} &= \sum_{m=0}^{N-1} h_1[m]h_2[m+n] \\
\mathbf{R}_{1,3} &= \sum_{m=0}^{N-1} h_1[m]h_3[m+n] \\
&\vdots \\
\mathbf{R}_{1,M} &= \sum_{m=0}^{N-1} h_1[m]h_M[m+n] \\
&\vdots \\
\mathbf{R}_{M,M-1} &= \sum_{m=0}^{N-1} h_M[m]h_{M-1}[m+n]
\end{aligned}
\tag{3.6}
$$

where h_i and h_j are the windowed RIR and $\mathbf{R}_{i,j}$ is a vector containing the cross-correlation function of microphones i and j. The argument of the maxima of the cross-correlation function determines the relative delay $\tau_{i,j}$ between each microphone pair.

$$
\begin{aligned}
\tau_{1,2} &= \arg\max\{\mathbf{R}_{1,2}\} = (\mathbf{m}_1 - \mathbf{m}_2)\frac{\mathbf{n}^T}{c} \\
\tau_{1,3} &= \arg\max\{\mathbf{R}_{1,3}\} = (\mathbf{m}_1 - \mathbf{m}_3)\frac{\mathbf{n}^T}{c} \\
&\vdots \\
\tau_{1,M} &= \arg\max\{\mathbf{R}_{1,M}\} = (\mathbf{m}_1 - \mathbf{m}_3)\frac{\mathbf{n}^T}{c} \\
&\vdots \\
\tau_{M,M-1} &= \arg\max\{\mathbf{R}_{M,M-1}\} = (\mathbf{m}_M - \mathbf{m}_{M-1})\frac{\mathbf{n}^T}{c}
\end{aligned}
\tag{3.7}
$$

The delay $\tau_{i,j}$ is related to the position $\mathbf{m}$ of every microphone and the direction of incidence $\mathbf{n}$ of the sound event. The vectors $\mathbf{m}$ and $\mathbf{n}$ represent Cartesian coordinates in the form

$[x, y, z]$. The previous expressions can then be expressed in matricial form to obtain a compact notation.

$$\mathbf{V} = \left[(\mathbf{m}_1 - \mathbf{m}_2)^T \quad (\mathbf{m}_1 - \mathbf{m}_3)^T \quad (\mathbf{m}_1 - \mathbf{m}_M)^T \quad (\mathbf{m}_M - \mathbf{m}_{M-1})^T\right]^T \tag{3.8}$$

$$\boldsymbol{\tau} = \left[\tau_{1,2} \quad \tau_{1,3} \quad \dots \quad \tau_{1,M} \quad \dots \quad \tau_{M,M-1}\right] \tag{3.9}$$

$$\boldsymbol{\tau} c = \mathbf{V}\mathbf{n} \tag{3.10}$$

Finally, the direction of incidence $\mathbf{n}$ is computed by calculating the Moore-Penrose pseudoinverse of the matrix of relative distances between microphone pairs $\mathbf{V}$, leading to the minimum mean square error solution.

$$\mathbf{n} = \mathbf{V}^{+}\boldsymbol{\tau} c \tag{3.11}$$

In practice, the accuracy of the localization procedure depends on a correct estimation of the relative delays between microphones. This means that the early part of the impulse response, where the echo density is lower, can be analyzed with higher accuracy. Furthermore, increasing the number of microphones adds robustness to the localization. After the localization procedure is completed the time window is shifted one sample and the process is repeated until the entire MRIR is analyzed.

SDM has been applied to numerous studies related to perceptual evaluation of concert hall acoustics [PL16], night clubs [Ter+15b], car cabin acoustics [Ter+15a] or studio control rooms [Ter+14].

3.1.3 Vector Base Amplitude Panning

Vector Base Amplitude Panning (VBAP) [Pul97] is a three-dimensional generalization of the two-channel stereophonic reproduction method. Three loudspeakers defining a triangular plane reproduce sound simultaneously with different gain, creating the sensory illusion of a phantom source at a specific location inside this plane. By means of a spherical loudspeaker set-up it is then possible to create a phantom source at any position. Given that the illusion of a phantom source is product of coherent summation of signals, it is generally necessary to place the loudspeakers equidistant from the listening position. In addition, the coherent summation is present only at the center of the sphere, being it the sweet spot. The layout of the reproduction set-up can, in principle, be arbitrary, although the distances between loudspeakers (size of triangles) have an effect on the localization sharpness of the virtual sources.

Once a loudspeaker set-up is defined, the first step is to generate an appropriate set of loudspeaker base triplets. The approach used in this work is a Delaunay triangulation of the loudspeaker positions followed by the extraction of the convex hull. Once the triplets are determined, and knowing the positions of all loudspeakers and the location of a desired virtual source, it is straightforward to determine which loudspeaker triplet should be active for the reproduction.

The sum of the power factors from each active loudspeaker must satisfy a constant sound power value C, regardless of the location of the virtual source:

$$g_1^2 + g_2^2 + g_3^2 = C \tag{3.12}$$

Then, the gain factor of each loudspeaker depends uniquely on the relative position between every loudspeaker and the virtual source. Being $\mathbf{p} = [p_x \quad p_y \quad p_z]^T$ and $\mathbf{l_i} = [l_{ix} \quad l_{iy} \quad l_{iz}]^T$ unit vectors containing the cartesian coordinates of the virtual source and a loudspeaker i, respectively, a linear combination of three loudspeaker vectors can express de position of the virtual source.

$$\mathbf{p} = g_1\mathbf{l}_1 + g_2\mathbf{l}_2 + g_3\mathbf{l}_3 \tag{3.13}$$

The previous expression can be written in matricial form by defining a vector of gain factors $\mathbf{g} = [g_1 \quad g_2 \quad g_3]$ and a matrix containing the positions of a loudspeaker triplet $\mathbf{L} = [\mathbf{l}_1 \quad \mathbf{l}_2 \quad \mathbf{l}_3]^T$.

$$\mathbf{p}^T = \mathbf{gL} \tag{3.14}$$

The system is solved by inversion of $\mathbf{L}$, which exists if the basis defined by L conforms a three-dimensional space.

$$\mathbf{g} = \mathbf{p}^T\mathbf{L}^{-1} = \begin{bmatrix} p_x & p_y & p_z \end{bmatrix} \begin{bmatrix} l_{1x} & l_{1y} & l_{1z} \\ l_{2x} & l_{2y} & l_{2z} \\ l_{3x} & l_{3y} & l_{3z} \end{bmatrix}^{-1} \tag{3.15}$$

VBAP represents an efficient and flexible method to map virtual sound sources in a three-dimensional space. The localization of panned sounds was studied by the creator of VBAP, Ville Pulkki in [Pul01]. The creation of a virtual source is product of a coherent summation of multiple signals and the localization cues used by the auditory system differ for lateral or elevated sounds. While horizontal localization depends mostly on Interaural Time Differences (ITD) and Interaural Level Differences (ILD) [Bla96], the localization of elevated sounds relies on spectral cues [Bla69]. As a result, the localization accuracy of amplitude panned sources depends on the location of the virtual sources. Sources panned close to the horizontal plane are decoded quite accurately, while elevated sources usually result on the identification of individual loudspeakers with different elevation. Therefore, a loudspeaker layout with higher vertical than horizontal density of loudspeakers can contribute to decrease these effects when using VBAP rendering.

3.1.4 Real-Time Convolution

Convolution is a central operation in any auralization process, combining the response of a space with a live or recorded sound. The formal definition of a discrete convolution $*$ is:

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k] \tag{3.16}$$

where x is the input signal, h is the impulse response and y is the output signal. The convolution operation can be performed also in the frequency domain, being it a multiplication. The implementation of the convolution in the time domain operation is straightforward and can in theory be implemented in real time. In practice, the computational cost of the time domain convolution is on the order of N^2 operations, and a real-time implementation is impracticable. The computational cost of convolution in the frequency domain is reduced to $Nlog(N)$, thanks to the efficiency of the Fast Fourier Transform (FFT) algorithm [CT65].

$$y[n] = IDFT\{DFT\{h[n]\} \cdot DFT\{x[n]\}\} \tag{3.17}$$

where $DFT\{\}$ and $IDFT\{\}$ denote the discrete fourier transform and inverse discrete fourier transform operations.

The implementation of convolution in the frequency domain requires the processing of a block of samples, meaning that it cannot be implemented in real-time. In addition, convolution in the frequency domain is cyclic, while in the time domain it is linear. To convolve a continuous input signal it is necessary to split the input signal into blocks, perform a separate convolution of each block and finally combine the results of successive convolved blocks appropriately in order to create the output signal. The most common methods for performing cyclic block based convolution in the frequency domain are overlap-save and overlap-add [Bur08].

A solution for a practicable real-time implementation is a fixed partition scheme, which consists on splitting the impulse response $h[n]$ into several blocks and performing convolution in the time domain for the early portion of the impulse response and the frequency domain convolution for the rest of the blocks [Gar95]. This hybrid approach allows an efficient implementation without delays. The tool used for real-time convolution in this project is the ~multiconvolve object for Max/MSP, part of the HISSTools externals package [HT12], and is based on the presented fixed partition scheme. In practice, since a live signal needs to be digitalized to perform a discrete convolution, the overall input/output delay is determined by the size of the audio buffers used in the AD/DA conversion.

3.2 Music Performance Analysis (MPA)

3.2.1 MIDI analysis

The extraction of performance related information from a MIDI recording is straightforward in the sense that mechanical actions performed by a musician are encoded in a MIDI stream. The most relevant information for MPA of key instruments, such as piano or organ, are pitch, note onset, note offset, velocity and pedal onset and offset.

The analysis of MIDI recordings through this work is focused on organ performance, and since the dynamic characteristics of the produced sound can not be modified directly by the player, only temporal parameters are studied. Although a performer is able to modify the active registers of the organ, these changes are usually not produced during the course of a musical piece, and they are not recorded in a MIDI file.

The analysis of temporal characteristics in MIDI recordings profits greatly from the availability of the music score. This allows a direct comparison between the recorded performance and the ground truth piece (score). The MIDI Toolbox for Matlab [ET04; TE16] is used to import MIDI files into Matlab, extract the basic note information of recordings and visualize recorded performances. The imported MIDI recordings and scores are expressed in NMAT format, which contains relevant information for every note: the MIDI channel, pitch, velocity and onset and duration times (both in beats and seconds). An example of the imported data and visualization is provided in Fig. 3.4. During this thesis, a set of functions have been completed to extract musical features from MIDI recordings. These functions focus on the evaluation of temporal aspects of the musical performance.

The duration of a performance T_{time} can be defined as the time between the first onset t_1 and the last onset t_{last}. Knowing the duration of the performance and the number of beats N_{beats}, the average tempo T_{tempo} in beats per minute (BPM) can be computed:

$$T_{tot} = t_{last} - t_1 \tag{3.18}$$

$$TMP_{tot} = 60 \cdot \frac{N_{beats}}{T_{time}} \tag{3.19}$$

Onset (beats)	Duration (beats)	MIDI channel	MIDI pitch	Velocity	Onset (sec)	Duration (sec)
0.0000	1.2250	7.00	48.00	127.00	0.0000	0.6125
0.0333	0.6354	1.00	63.00	127.00	0.0167	0.3177
0.0490	0.9073	1.00	72.00	127.00	0.0245	0.4536
0.0781	2.0750	1.00	67.00	127.00	0.0391	1.0375
0.1063	1.9458	1.00	55.00	127.00	0.0531	0.9729
1.0240	1.1542	7.00	50.00	127.00	0.5120	0.5771
1.0281	1.0958	1.00	65.00	127.00	0.5141	0.5479
1.0427	1.1396	1.00	71.00	127.00	0.5214	0.5698
3.1719	1.0948	7.00	48.00	127.00	1.5859	0.5474
3.2031	0.4490	1.00	63.00	127.00	1.6016	0.2245
3.2469	1.7719	1.00	55.00	127.00	1.6234	0.8859
3.2490	0.8198	1.00	72.00	127.00	1.6245	0.4099
⋮	⋮	⋮	⋮	⋮	⋮	⋮

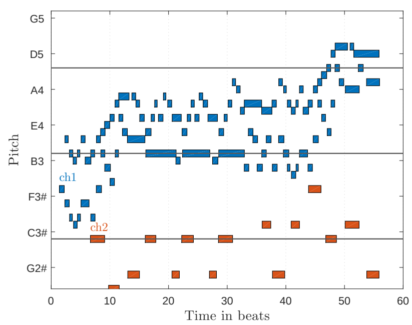


Fig. 3.4.: Decoded data stream (left) and pianoroll visualization (right) of a MIDI organ recording.

If the MIDI score is available arrays $TMP_{note}(i)$ and $TMP_{bar}(i)$ containing the inter-note tempo and inter-bar tempo, respectively, can be expressed as

$$TMP_{note}(i) = 60 \cdot \frac{b_{on}(i+1) - b_{on}(i)}{t_{on}(i+1) - t_{on}(i)} \tag{3.20}$$

$$TMP_{bar}(i) = 60 \cdot \frac{N_{meas}}{tbar_{on}(i+1) - tbar_{on}(i)} \tag{3.21}$$

where $b_{on}(i)$ and $t_{on}(i)$ are the onset beat and onset time of the note i, respectively, $tbar_{on}(i)$ is the onset time of the first note of the bar i, and N_{meas} is the number of beats per bar.

The acoustic feedback of a room and effects of reverberation are most heard after note offsets. The computation of average break duration can provide insightful data on a player's reaction to acoustics.

$$T_{\mu break}(s) = \frac{1}{N}\sum_{i=1}^{N} t_{rest}(i) \tag{3.22}$$

3.2.2 Audio analysis

Since audio recordings do not contain explicit information about performance characteristics, the typical approach to the problem consists on automatic extraction of a number of audio features. These features usually provide information about different domains of the recording, providing e.g. temporal, spectral, timbral or energetic information. Most of the audio features that are extracted from music recordings during this work are available in the MIR Toolbox for Matlab [LT07]. The extraction of short and long term signal envelopes is based on the CUEX algorithms [Fri+07]. The estimation of temporal curves is based on a combination of temporal alignment of recordings with a ground-truth annotated audio file, using Dynamic Time Warping (DTW) and inspired by the work presented by Turetsky and Ellis in [TE03]. A schematic view of the extracted features is displayed in Fig. 3.5.

A short description of the features is provided below, with a reference to the calculation method:

- *rms*: Root-mean-square (RMS) value of the waveform. Computed using the command *mirrms* from the MIR Toolbox [LT07].

- *rmsA*: Root-mean-square value of the waveform filtered using an A-weighting filter. Computed using the command *mirrms()* from the MIR Toolbox [LT07].

- *LUFSlinear*: Average value (in linear scale) of the Gated Loudness Units relative to Full Scale, computed according to ITU 1770-4 [Uni15].

- *LUFSstd*: Standard deviation (in linear scale) of the Gated Loudness Units relative to Full Scale, computed according to ITU 1770-4 [Uni15].

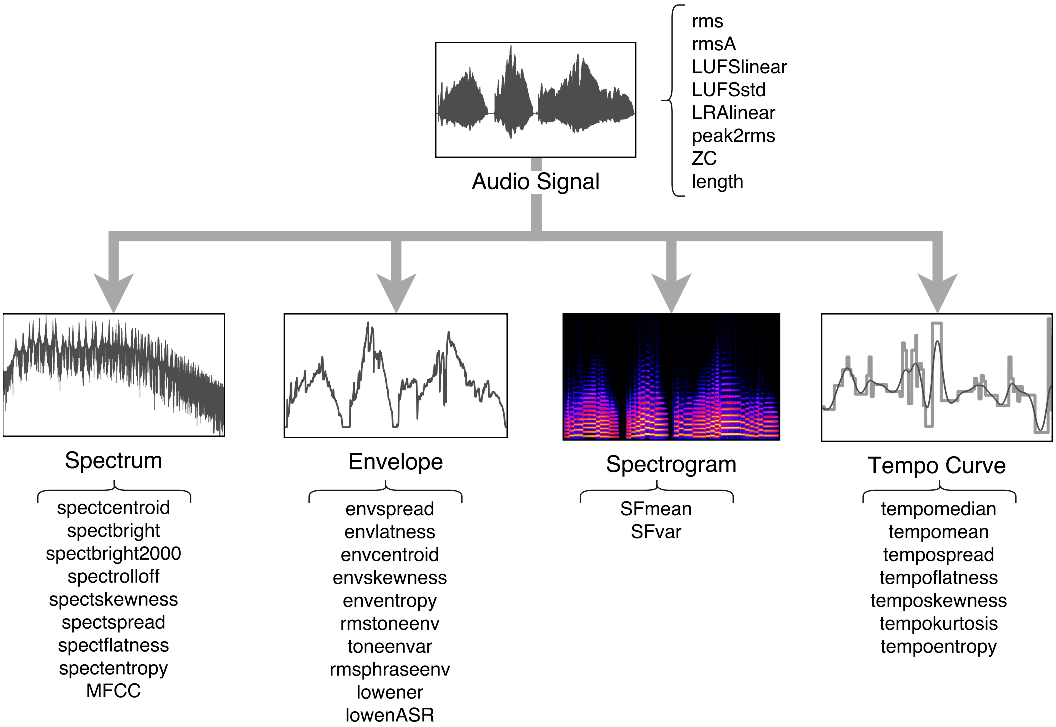


Fig. 3.5.: Audio features extracted from the audio signals.

- *SFmean*: Average value of the spectral flux of the audio signal. COmputed using the *mirflux()* command from the MIR Toolbox [LT07]

- *SFvar*: Standard deviation of the spectral flux of the audio signal. COmputed using the *mirflux()* command from the MIR Toolbox [LT07]

- *LRAlinear*: Loudness range (in linear scale) of the audio signal, computed according to EBU Tech 3342 [Uni16].

- *peak2rms*: Ratio of largest absolute value of the waveform by the RMS value, computed using the *peak2rms()* built-in Matlab function.

- *envspread*: Standard deviation of the temporal envelope computed using the *mirenvelope()* command from the MIR. Toolbox [LT07].

- *envflatness*: Flatness of the temporal envelope (*mirenvelope()*) computed using the *mirflatness()* command from the MIR Toolbox [LT07]. Flatness is defined as the ratio between the geometric mean and the arithmetic mean of a data series.

- *envcentroid*: Centroid of the envelope function of the waveform. Computed using the *mircentroid()* command from the MIR Toolbox [LT07].

- *envskewness*: Centroid of the envelope function of the waveform. Computed using the *mirskewness()* command from the MIR Toolbox [LT07].

- *enventropy*: Entropy of the envelope function of the waveform. Computed using the *mirentropy()* command from the MIR Toolbox [LT07].
- *rmstoneenv*: RMS value of the tone envelope function. The tone envelope is computed according to the CUEX algorithms [Fri+07].
- *toneenvar*: Standard deviation of the tone envelope function. The tone envelope is computed according to the CUEX algorithms [Fri+07].
- *rmsphraseenv*: RMS value of the phrase envelope function. The phrase envelope is computed according to the CUEX algorithms [Fri+07].
- *lowener*: Low energy rate, defined as the ratio between the instant energy value of signal frames over the average frame energy. Computed using the *mirlowenergy()* command from the MIR Toolbox [LT07].
- *lowenASR*: Average-to-silence ratio, computed as the ratio between number of frames with less energy than an arbitrary threshold and the number of frames of an audio signal. Computed with the *mirlowenergy(,'ASR')* command from the MIR Toolbox [LT07].
- *spectcentroid*: Centroid of the spectrum of an audio signal. Computed using the *mircentroid()* command from the MIR Toolbox [LT07].
- *spectbright*: Spectral brightess, defined as the amount of energy above 1500 Hz. Computed using the *mirbrightness()* command from the MIR Toolbox [LT07].
- *spectbright2000*: Spectral brightess with a threshold of 2000 Hz, defined as the amount of energy above 2000 Hz. Computed using the *mirbrightness()* command from the MIR Toolbox [LT07].
- *spectrolloff*: Spectral rolloff, defined as the frequency below of which the 85% of the energy is found. Computed using the *mirrolloff()* command from the MIR Toolbox [LT07].
- *spectskewness*: Spectral skewness, defined as the coefficient of skewness of the spectrum of the audio signal. Computed using the *mirskewness()* command from the MIR Toolbox [LT07].
- *spectspread*: Spectral spread, defined as the standard deviation of the spectrum of the audio signal. Computed using the *mirspread()* command from the MIR Toolbox [LT07].
- *spectflatness*: Flatness of the spectrum of the signal computed using the *mirflatness()* command from the MIR Toolbox [LT07].
- *spectentropy*: Entropy of the spectrum of the signal, computed using the *mirentropy()* command from the MIR Toolbox [LT07].

- *MFCC1* to *MFCC9*: Mel-Frequency Cepstral Coefficients 1 to 9, computed using the *mirmfcc()* command from the MIR Toolbox [LT07].

- *ZC*: Zero Cross, or total number of zero crossings of the signal, computed with the *mirzerocross()* command from the MIR Toolbox [LT07].

- *length*: Total length of the audio signal, from the first onset to the last offset.

- *tempomedian*: Median value of the *tempo* curve of the signal.

- *tempomean*: Average value of the *tempo* curve of the signal.

- *tempospread*: Standard deviation of the *tempo* curve of the signal.

- *tempoflatness*: Flatness of the *tempo* curve of the signal.

- *temposkewness*: Skewness coefficient of the *tempo* curve of the signal.

- *tempokurtosis*: Kurtosis of the *tempo* curve of the signal.

- *tempoentropy*: Entropy of the *tempo* curve of the signal.

Extraction of *tempo* curves

To compute temporal features - *tempomedian, tempomean, tempospread, tempoflatness, temposkewness, tempokurtosis, and tempoentropy* - it is necessary to estimate a *tempo* curve, from which the parameters are then calculated. An inter-note *tempo* curve represents the instantaneous *tempo* of a musical recording at the beginning of each note, and it is calculated as in Eq. 3.21. Thus, for the estimation of *tempo* curves it is necessary to have the onset time and beat number of every note in the recording. While the beat number is easily obtained from a musical score, several approaches can be followed to estimate note onsets, as explained in Sec. 2.3.1. Since the accuracy of automatic extraction of onset times is often insufficient, and manual annotation of onset times is laborious and impracticable with large datasets of recordings, an alternative approach combining both methods is proposed and used here.

The main idea behind the method consists on manually annotating only one of the recordings of each musical piece in the dataset. Then, using Dynamic Time Warping (DTW) [Ell03; TE03], another recording of the same piece is aligned with the reference recording and the target time indices on the alignment path are compared to estimate the onsets times of the second recording.

The detailed process is as follows:

1. A MIDI score of the piece is generated.

2. Onset times for a reference audio recording are manually annotated.

3. A reference *tempo* curve is generated comparing the score onsets with the reference recording onsets.

4. Dynamic Time Warping (DTW) is applied on the reference and target audio recordings, obtaining an alignment path.

5. Onset times of the target recording are obtained by comparing the desired reference and target time indices on the alignment path.

6. A tempo curve is extracted comparing the target onset times and the MIDI score onset times.

7. The tempo curve is resampled to regular sampling times.

8. The tempo curve is low pass filtered.

Although the described process uses previously available tools to align audio recordings – mainly the DTW algorithms by Ellis [Ell03] –, the *tempo* curve estimation process as described in this section constitutes a new approach to improve the accuracy of (semi)automatic onset annotation. Previous works on music performance analysis used either manual or automatic onset annotation [SK15; Kat+15; Fri+07]. Additionally, audio-to-score alignment is a task that frequently takes advantage of DTW [CO+15; TE03; Mül+09; Kon+09] by aligning audio recordings to ground-truth MIDI scores. However, to the best knowledge of the author, up to the current moment there has been no work combining audio-to-midi and audio-to-audio alignment to extract *tempo* curves, thus being the proposed method a new approach to complete this task. The main advantage of this alternative is that audio-to-audio alignment yields better perceptual results than audio-to-MIDI alignment when applied to the recordings of this project. Then, having only one reference audio recording with annotated onsets allows the extraction of *tempo* curves of other recordings of the same piece.

Note that the estimated *tempo* curves are generated from perceptual onset times [VR81], given that the onset times of the reference audio recording are annotated a human listener. The method was informally evaluated by human listening to several trumpet recordings, and there are no perceptually relevant differences between the reference and the aligned audio signals, suggesting that the accuracy of the method is sufficient for the application to music signals.

4 Implementation of a virtual acoustic environment - Detmold Surround Sound Sphere (D3S)

A virtual acoustic environment allows users to be immersed in real or simulated acoustic scenes by means of an electro-acoustic setup and appropriate digital signal processing. This chapter focuses on the implementation of a virtual environment that intends to replicate the acoustic conditions of real performance spaces present in the University of Music Detmold, allowing real-time interaction between musicians and virtual rooms. This means that a musician is presented with the appropriate acoustic feedback of different spaces as if they were playing in the real space.

The goal of the virtual environment presented in this thesis is to realistically reproduce the acoustics of a real space in real-time. This can be achieved by reproducing every reflection correctly in terms of timing, direction, amplitude and spectrum. The first step in the auralization process is the measurement of spatial room impulse responses (SRIR) using a directional source and a microphone array on stage. After that, the captured SRIR are analyzed using SDM and resynthesized, generating appropriate filters for a specific loudspeaker setup. The sound of a musician is captured using a directional microphone and convolved in real time with the filters, resulting in the acoustic feedback of a specific space. The following sections describe in detail the measured rooms, technical setup, signal processing operations and validation of the system. The process presented in this chapter is described with a particular application to auralization of rooms for trumpet players.

4.1 Acoustic measurements

4.1.1 Measurement setup

The measurement setup is composed by two main elements: a directional sound source and a microphone array. Besides that, the setup contains appropriate AD/DA converters, microphone preamplifier, sound interface and a laptop.

In order to achieve a realistic auralization it is desirable to excite the room in a similar manner as a real instrument would, specially in terms of radiation. For this reason, in this thesis a studio monitor (Neumann KH120 A) is used, since its radiation pattern presents great similarities to a trumpet (see Fig 4.1). The directivity measurements were conducted in an anechoic room using a circular microphone array composed of 24 omnidirectional measurement microphones (Beyerdynamic MM1). While the loudspeaker radiation measurements were part of this project, the trumpet directivity data was kindly provided by Dr. Grothe, and corresponds to a seated player and a trumpet (unpublished work at the time of writing this document).

As can be observed in Fig 4.1, the directivity difference between loudspeaker and trumpet is contained within 5 dB in the range 250 - 4000 Hz. At higher frequencies the trumpet presents a greater directivity towards the front, presenting a set of secondary lobes in the range 8 - 16 kHz. Although the spectral content at high frequencies is significant for high playing levels [Luc75], at normal levels most of the energy of a trumpet is concentrated below 4000 Hz and the spectral envelope level at 4000 Hz is approximately 30 dB less than the maximum level, with a decay of 15 dB/octave [LC67]. Thus, this approach can be considered sufficiently accurate for the given purposes as it provides a good spatial match at the bands with higher energy. Since the main analysis and synthesis method used in

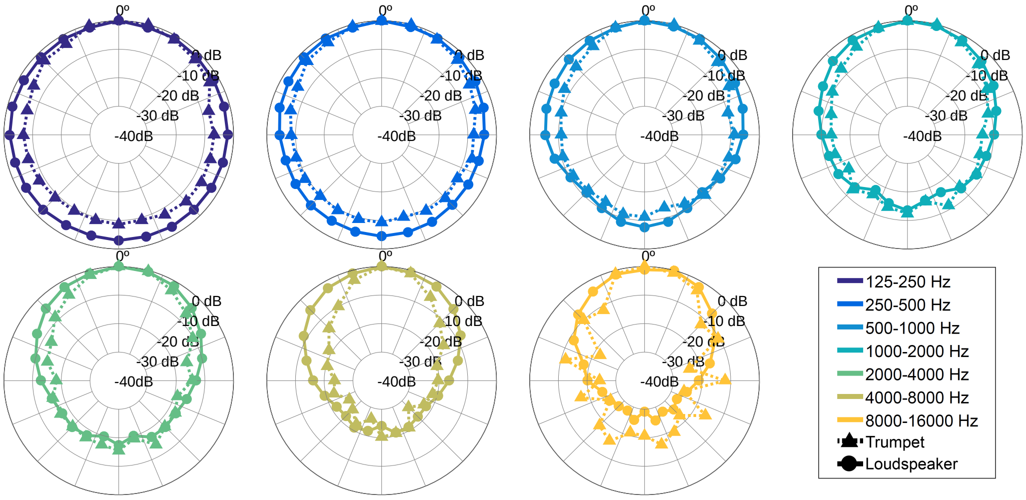


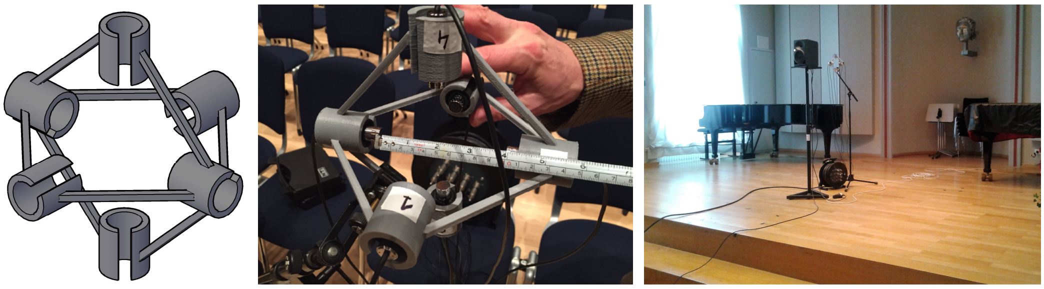
Fig. 4.1.: Radiation patterns (horizontal plane) of a studio monitor (solid lines, circular markers) and a trumpet (dotted lines, triangle markers).

this project is SDM, the design of the microphone array was inspired by previous research, following the guidelines of Tervo *et al.* in [Ter+13]. The microphone array is an open array composed of 6 omnidirectional measurement microphones (NTi 2010M) arranged on a 3D space in orthogonal directions. The spacing between opposed microphones is 10 cm and their axis are directed towards the center of the array. The array holder was designed during this project and manufactured with a 3D printer.

The acoustic source and microphone array were arranged on stage imitating the configuration of a trumpet player performing on stage. The height of the source was approximately 1.5 m above the stage, and the microphone array was positioned slightly higher at approximately 60 cm, replicating the positions of mouth and ears of a musician (see Fig. 4.2).

4.1.2 Rooms description

The auralization method implemented in this thesis is based on the resynthesis of existing rooms. For this reason, three rooms from the University of Music Detmold were measured using spatial techniques (see Fig. 4.3):



(a) (b) (c)

Fig. 4.2.: a) CAD model of the microphone array holder prototype; b) Microphone array; c) Measurement setup on stage.

Room	Abbreviation	Description	Seats	Volume (m^3)	Stage vol. (m^3)
Brahmssaal	BS	Small performance room	110	750	230
Detmold Konzerthaus	KH	Medium sized concert hall	600	4600	600
Detmold Sommertheater	DST	Small theater	320	2700	650

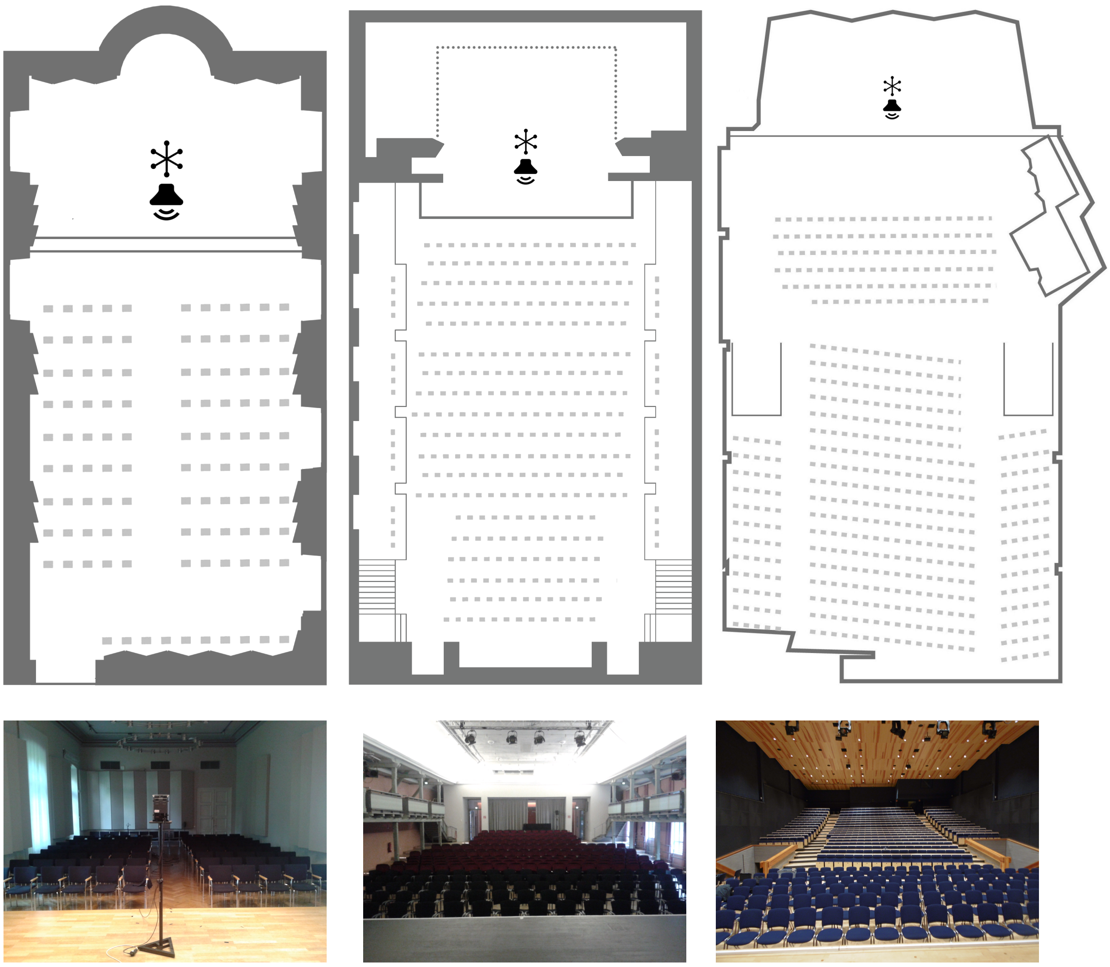
Tab. 4.1.: General information of the measured rooms.

- **Brahmssaal (BS)** is a small performance room with capacity for approximately 110 persons and shoe-box shape. The seating arrangement is composed of removable seats with covered by a thin layer of foam. The floor is made of laminated wood and walls and ceiling are painted concrete. The lateral and back walls are covered with several sets of sound reflectors. The room is symmetrical except for the door position and a few lateral windows. The typical uses of the room are individual instrumental lessons, solo performance and small ensemble concerts.

- **Detmold Sommertheater (DST)** is a theater with capacity for approximately 320 persons and shoe-box shape. The audience is distributed on a main floor and elevated side balconies. The main floor is furnished with a first section of removable chairs and a larger area consisting of upholstered seats with heavy absorbing. The main floor is made of laminated wood, while walls and ceiling are made of painted concrete. The theater includes a coverable orchestra pit and a fly tower. The walls of the stage are covered with acoustic absorbing curtains. The main uses of the room are theater, opera, solo performances, small ensembles and amplified music.

- **Detmold Konzerthaus (KH)** is the biggest performance room present in the university. It has capacity for approximately 600 listeners and is equipped with a pipe organ, a surrounding loudspeaker setup and a coverable orchestra pit. The floor plan presents a shoe-boxlike shape and the walls material is painted concrete. The audience floor is inclined and the ceiling is composed of panels with different inclinations. The floor and ceiling material is laminated wood. A portion of the audience area (in front of the stage) is furnished with removable seats and it can be cleared for artistic performances or recordings. The rest of the audience area is furnished with upholstered seats. The room is commonly used for symphonic and chamber orchestra performances, as well as organ recitals, solo concert examinations and music productions.

A summary with the key information of every room is included in Table 4.1.



(a) BS (b) DST (c) KH

Fig. 4.3.: Floor plans and general view of the measured rooms.

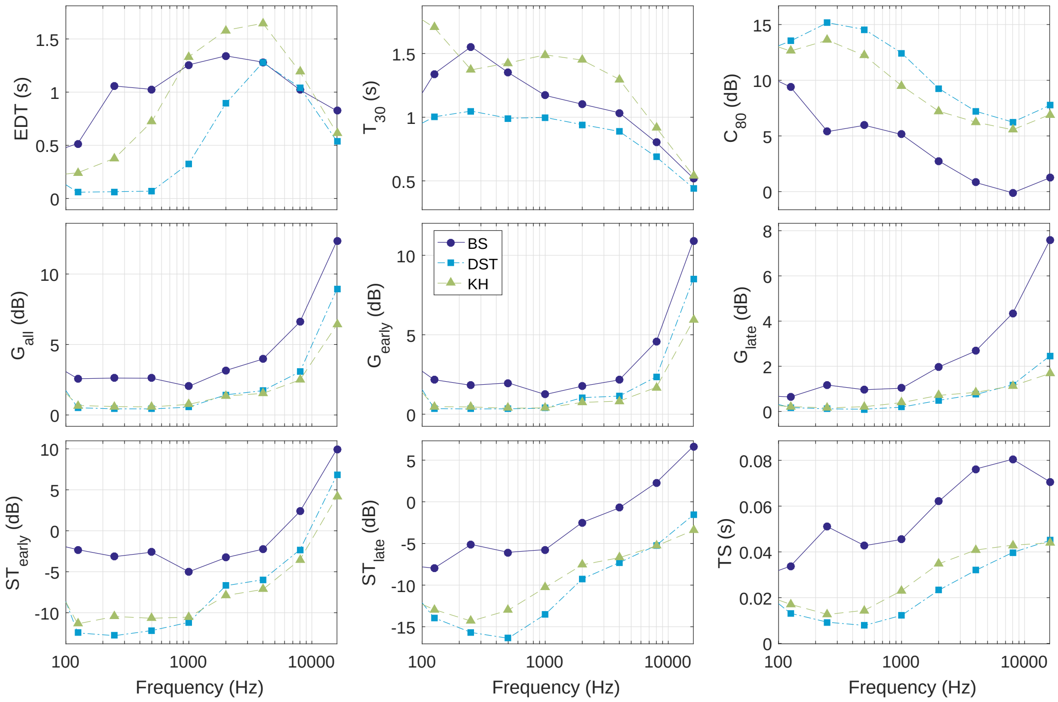


Fig. 4.4.: Monaural room acoustic parameters measured on stage.

4.1.3 Room acoustical properties

Using measured stage IR the following monaural parameters have been calculated: EDT, T_{30}, C_{80}, G_{all}, G_{early}, G_{late}, ST_{early}, ST_{late}, G_{early}, G_{late}, and Center Time (see Fig. 4.4). The parameters have been derived in octave bands for every channel of the SRIR and averaged over all channels afterwards. In addition, room strength (G) and stage parameters (ST) are measured relative to the direct sound instead of their usual definitions. This deviation from the standard responds to the fact that the measurement setup differs importantly from the procedure described in the standard ISO 3382-1 [ISO09]. However, the obtained values describe a closer representation of the actual energy ratios perceived by the musician at the playing position.

Using the measured SRIR, spatial information of the room has been analyzed using SDM and spatiotemporal plots of the rooms were then generated [Pät+13] (see Fig. 4.5).

4.2 Signal processing

The signal processing operations can be divided into two main categories: off-line operations and online - or real-time - operations. The first category is related with the measurement and analysis of SRIR, postprocessing operations and generation of appropriate filters for later convolution. The real-time operations deal with the convolution and equalization operations of the previously derived filters.

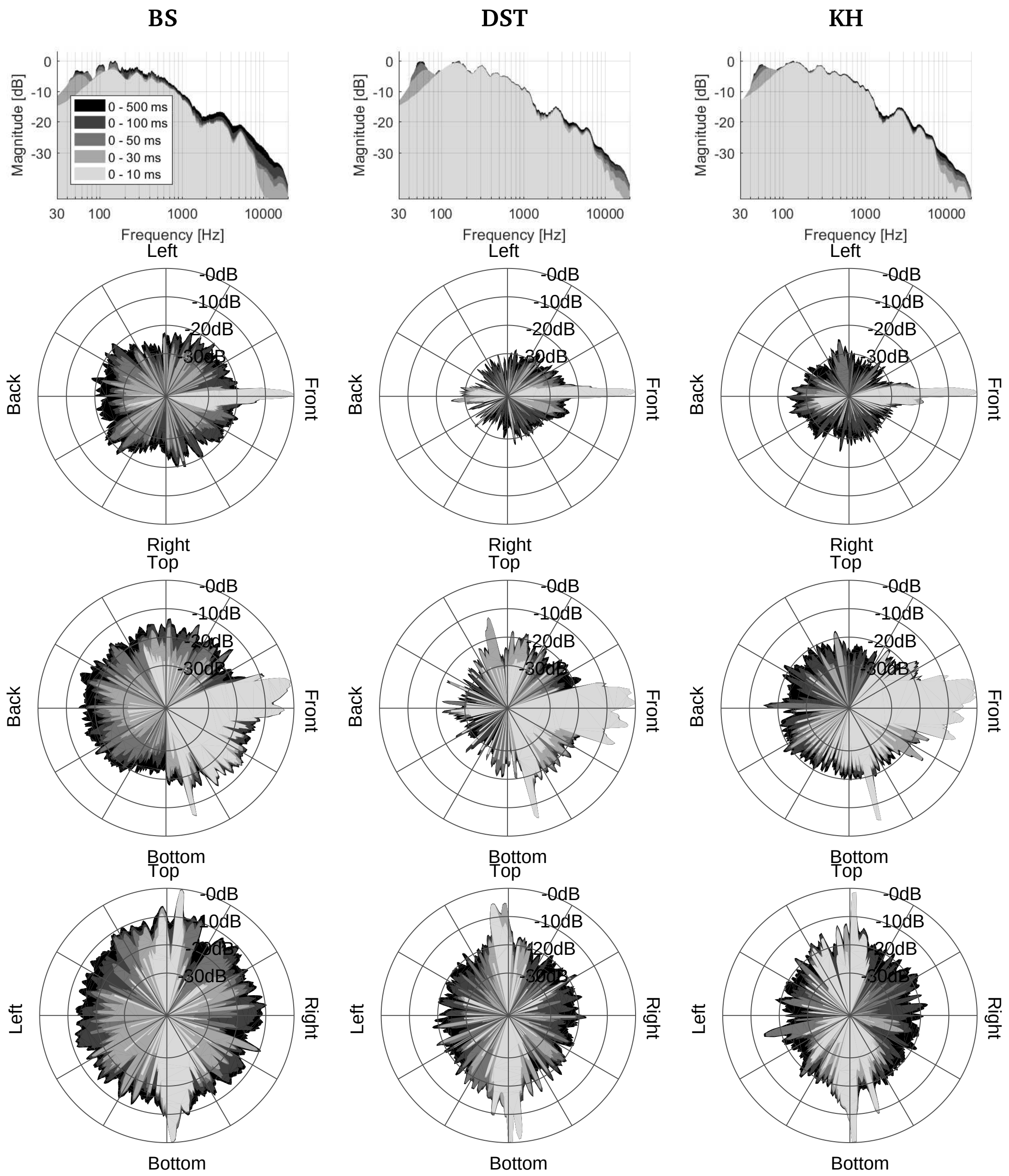


Fig. 4.5.: Time-frequency and spatio-temporal representations of the measured rooms.

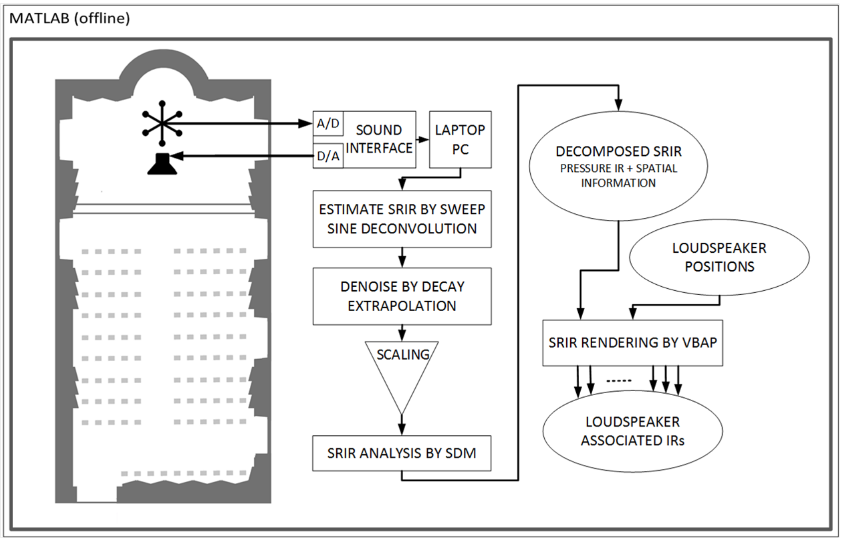


Fig. 4.6.: Room analysis and resynthesis operations performed in the D3S.

4.2.1 Spatial analysis and resynthesis

A measurement setup is arranged on stage imitating the configuration of a musician and their instrument. Spatial Room Impulse Responses are obtained using the sweep deconvolution technique [Far00]. They are then denoised extrapolating the energy decay of the impulse responses in octave bands. The direct sound of the impulse responses is normalized to ensure the same relative energy between different measured rooms. The spatial information of the SRIR is obtained using SDM, resulting into one vector of directions associated to every sample of a pressure impulse response. Finally, filters for every loudspeaker are obtained by mapping the pressure impulse response to the appropriate directions using VBAP. The filters are saved into multichannel audio files for later convolution. A block diagram showing the operations is depicted in Fig. 4.6. All these operations are performed offline and implemented in Matlab.

Manipulation of spatial information

The analysis of a SRIR with SDM results in a monaural RIR with associated directional information for every sample (parametrized SRIR). Thus, the directional information of the analyzed SRIR can be manipulated before resynthesizing them. This section describes the procedure applied to generate resynthesized SRIR with directional early energy.

A decomposed SRIR can be expressed as

$$IR(t, \theta, \phi) = SDM\{IR(t)_i)\} \tag{4.1}$$

where SDM refers to the spatial analysis of a multichannel SRIR with i channels.

A decomposed SRIR is split into two parts containing the modified early reflections and the original late reverberation, respectively. A weighting function is applied to the early reflections, and the two parts of the resulting SRIR are combined using a window function, representing the mixing time of the impulse response.

$$IR(t,\theta,\phi)_{dir} = IR(t,\theta,\phi)_{ERdir} + IR(t,\theta,\phi)_{LRorig} \tag{4.2}$$

$$IR(t,\theta,\phi)_{LRorig} = IR(t,\theta,\phi)_{orig} \cdot (1 - w(t)) \tag{4.3}$$

$$IR(t,\theta,\phi)_{ERdir} = IR(t,\theta,\phi)_{orig} \cdot g(\theta,\phi) \cdot w(t) \tag{4.4}$$

where $IR(t,\theta,\phi)_{dir}$ refers to a pressure impulse response with modified directional information. The window function $w(t)$ is used to separate the early reflections and late reverberation, $IR(t,\theta,\phi)_{ERdir}$ and $IR(t,\theta,\phi)_{LRorig}$, respectively. In addition, a directional weighting function $g(\theta,\phi)$ is applied to the early reflections. This results on a decomposed impulse response with modified early reflections, while the late reverberation remains intact.

The windowing and the weighting function will be determined depending on the application. In this work, five weighting functions are implemented in order to generate auralizations of the same rooms with different early energy directional properties. The generated auralizations are the original measured room (*all*), three responses with a figure of eight weighting on orthogonal directions (*front-back*, *sides*, *top-down*), and a complete removal of the early reflections (*no-ER*). The applied functions are as follows:

$$g(\theta,\phi) = \begin{cases} 1 & \text{if all} \\ |\cos(\theta)\cdot\cos(\phi)| & \text{if front-back} \\ |\sin(\theta)\cdot\cos(\phi)| & \text{if sides} \\ |\sin(\phi)| & \text{if top-down} \\ 0 & \text{if no-ER} \end{cases} \tag{4.5}$$

Finally, the time window $w(t)$ defines the transition between early and late reverberation and in this case is implemented as a linear cross-fade using two parameters, the start of the late reverberation, t_{end} and the mixing time t_{mix}.

$$w(t) = \begin{cases} 1 & \text{if } t < t_{end} - t_{mix} \\ -1/t_{mix} & \text{if } t_{end} - t_{mix} \leq t \leq t_{end} \\ 0 & \text{if } t > t_{end} \end{cases} \tag{4.6}$$

A schematic summary of the directional energy manipulation process is displayed in Fig. 4.7.

4.2.2 Real-time engine

The main requirement to implement an interactive system which allows musicians to perform in virtual rooms is the convolution in real-time of their sound and the resynthesized impulse responses. To achieve this, a real-time engine has been implemented using Max/MSP, whose main function is to perform real-time convolution. However, to improve the versatility

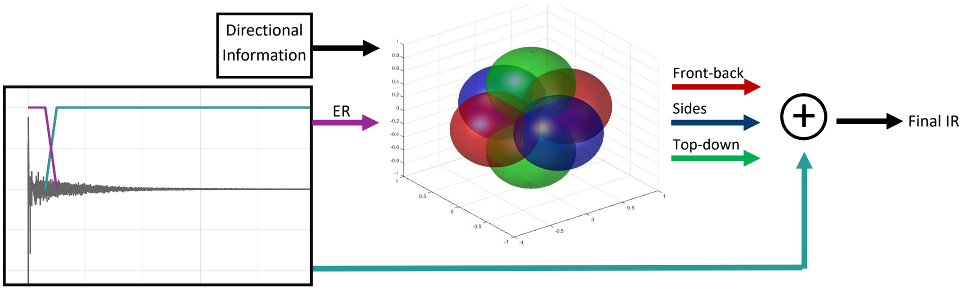


Fig. 4.7.: Manipulation process of early energy of analyzed SRIR.

Fig. 4.8.: GUI of the real-time engine of the D3S.

of the engine, complementary operations are implemented as well i.e. post-processing of resynthesized impulse responses, conditioning of input signal from the live musical instrument, individual gain controls on every output channel and estimation of the room acoustic parameters of the used impulse responses. A screenshot of the graphical user interface (GUI) is included in Fig. 4.8. A block diagram showing the main operations of the engine is shown in Fig. 4.9. The program reads the resynthesized room filters which are then saved into buffers. Since the musician is generating the direct sound in the virtual environment, the filters are time cropped and the direct sound is removed. In addition, depending on the used hardware i.e. sound interface, computer, AD/DA converters - and system configuration i.e. audio buffer size, convolution partition size - the round trip delay of the system will be different. For this reasons, the filters are further cropped, matching the time of arrival of the first reflections of the virtual environment and the measured SRIR.

The input signal is amplified to match the correct amplitude i.e. the amplitude of the direct sound. Furthermore, the frequency response of the system is affected by the response of the microphone and near-field artifacts due to the small distance between the instrument and the microphone. This is compensated for convolving a compensation filter (FIR) with the live sound of the instrument. The implementation of this filter is described in Section 4.3. Since the delay, amplification and equalization of the system depend on the used hardware setup,

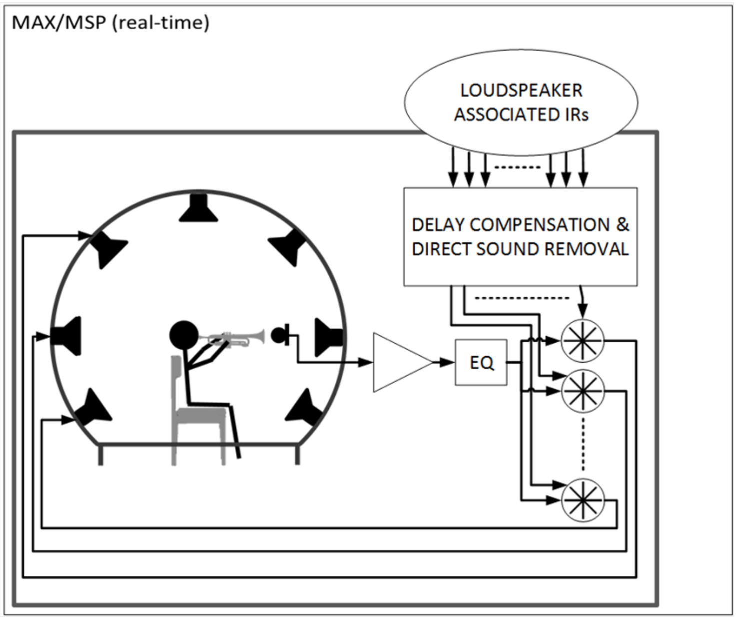


Fig. 4.9.: Main operations of the real-time engine of the D3S.

Buffer size (samples)	Computational delay (ms)	Total delay (ms)
64	5.9	8.7
128	8.6	11.4
256	14	16.8
512	24.6	27.4

Tab. 4.2.: Delay of the real-time engine of the D3S.

miking technique and instrument, the adjustment of the described parameters will vary significantly depending on the application and the parameters can be easily adjusted by the user. The specific calibration process applied to trumpet players is described in Section 4.3.

The real-time convolution is implemented using the ~multiconvolve object for Max/MSP from the HISSTools externals package [HT12]. This object performs zero-latency convolution of up to 64 input and output channels using a fixed partitioned scheme. To achieve this, the early part of the impulse response is convolved in the time domain, while the latter part implements FFT-based convolution. The implemented system features real-time convolution of up to 13 channels and can be scaled up easily. The round-trip delay of the real-time engine depends on the software and hardware configurations. In the experiments carried out during this thesis the round-trip delay from the microphone input to the listening position is 8.7 ms, from which the computational delay is 5.9 ms and the travelling time from the loudspeakers to the listening position is 2.8 ms. This is achieved using the zero delay option of the ~multiconvolve object and I/O audio buffer sizes of 64 samples (at 48 kHz). The delay has been measured using a hand-held device (NTi XL2), and other buffer sizes have been investigated. The round-trip delays for different configurations are shown in Tab. 4.2. The used impulse responses for convolution can be switched easily by using a pop-up menu or using Max/MSP messages, allowing the implementation of user interfaces for the conduction of experiments with listeners or musicians.

Frequency (Hz)	63	125	250	500	1000	2000	4000	8000	16000
RT_{30} (s)	0.41	0.23	0.10	0.09	0.08	0.06	0.05	0.04	0.04

Tab. 4.3.: Reverberation time of the listening room.

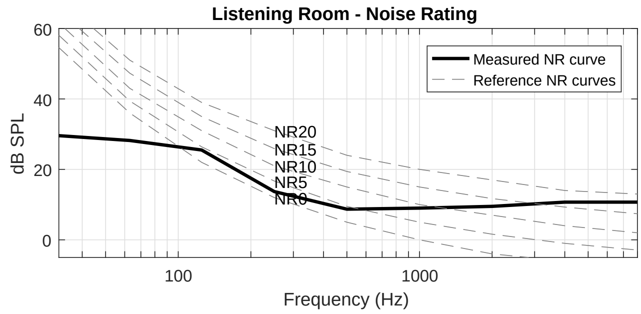


Fig. 4.10.: Noise Rating measurement of the listening room.

4.2.3 Reproduction setup

A loudspeaker based reproduction setup was implemented in a studio room of the Erich Thienhaus Institute. The inner dimensions of the room are $5.6 \times 4 \times 2.8$ meters. The room contains appropriate acoustic treatment which was complemented with broadband absorbing material in order to achieve quasi-anechoic conditions. The reverberation time RT_{30} of the room is less than 0.1 s at mid and high frequencies (Table 4.3 contains the reverberation time in octave bands). To derive the reverberation time monaural impulse responses from all the loudspeakers were obtained at the listening position and the results were then averaged. Noise rating (NR) measurements were performed using a hand-held calibrated sound level meter (NTi XL2). The noise floor of the room fulfills the NR19, which is fairly close to the recommended NR15 for listening rooms [Uni97] and complies with the ISO requirement of NR25 for concert halls and recording studios [ISO16].

A surrounding loudspeaker setup was designed and manufactured during this project to implement a mobile and flexible reproduction environment. The setup is composed of 13 active studio monitors (Neumann KH120 A) with flat response within ±2 dB from 54 Hz to 20 kHz. The loudspeakers are mounted on a rigid frame which can be disassembled for transport and arranged on three rings at elevation angles -35°, 0° and +45°. Every ring is composed of 4 loudspeakers placed every 90°. At 90° elevation there is a top loudspeaker. The positions of the loudspeakers can be easily modified in elevation, but they are kept constant during all the experiments of this project. The detailed positions of every channel are described in Tab. 4.4. A technical drawing of the loudspeaker arrangement are presented in Fig. 4.11.

Channel	1	2	3	4	5	6	7	8	9	10	11	12	13
Azimuth	225°	225°	225°	135°	135°	135	0°	45°	45°	45°	315°	315°	315°
Elevation	0°	-35°	45°	-35°	0°	45°	90°	0°	-35°	45°	0°	-35°	45°

Tab. 4.4.: Positions of the loudspeakers in the reproduction setup of the D3S.

Impulse response measurements of all the loudspeakers have been performed at the listening position to ensure a correct level and equalization. The frequency response of the loudspeak-

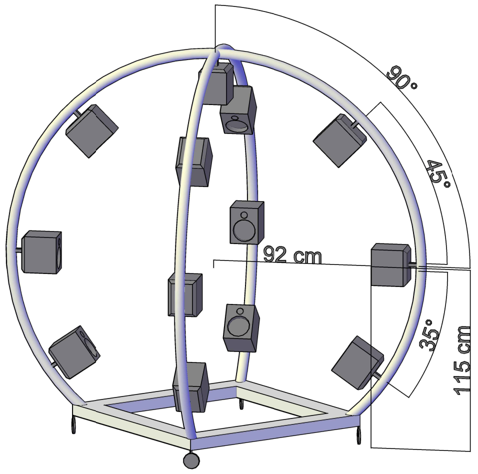
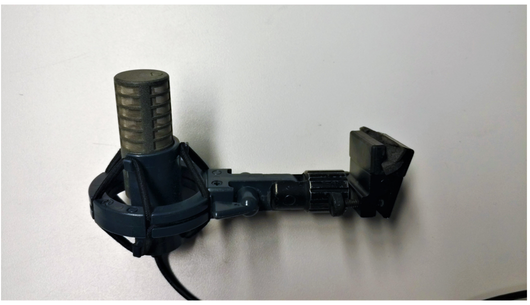
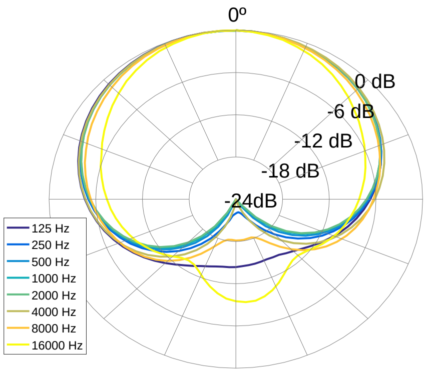


Fig. 4.11.: Technical drawing of the listening setup of the D3S.



(a)



(b)

Fig. 4.12.: Close view (a) and directivity (b) of the microphone Schoeps CCM 4V.

ers is predominantly flat. However, since the dimensions of the room are rather small some artifacts are present due to room modes, especially on channel 7 (top loudspeaker), which is close to the ceiling and situated a perpendicular line between the floor, the microphone (listening position) and the ceiling.

To capture the sound of an instrumentalist performing in a virtual environment a directional cardioid microphone (Schoeps CCM 4V) is used. The microphone is attached to the trumpet bell using a clip with the direction of maximum sensitivity pointed towards the instrument. The directivity allows a reduction of approximately 6 dB at 90° and 18 dB at the back between the bands 250 to 8000 Hz. This contributes to reducing significantly the input from the reproduced sounds to the microphone, ensuring the absence of feedback at usual playing levels. A close view of the microphone and its directivity are depicted in Fig. 4.12. The directivity has been measured in an anechoic room at a distance of 1 m using a turntable in steps of 5°. The frequency response of the microphone is predominantly flat and the possible deviations due to near-field effect are corrected in a calibration measurement described in Section 4.3.

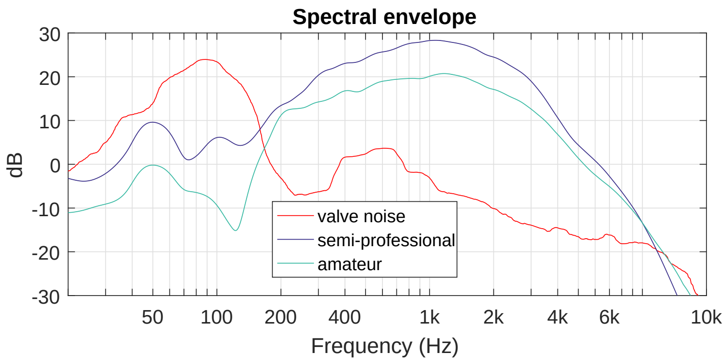


Fig. 4.13.: Spectral envelope of trumpet recordings and valve noise.

The audio reproduction chain is composed of a digital USB audio interface (RME Madiface XT), a computer running Max/MSP under operating system Windows 10 and a D/A converter SSL Alphalink SX. The microphone is directly connected to the audio interface, which is equipped with microphone inputs and phantom power supply. The outputs of the audio interface are connected to the D/A converter using a digital optical MADI connection. The signals from the D/A converter are then routed to the loudspeakers. In order to test the computational requirements of the implementation, two different computers were tested. Since both computers provided a satisfactory performance, they were used in the experiments depending on their availability. Technical details of the setup configuration and used equipment are provided in Appendix B. This information should allow a future implementation of the reproduction setup with calibrated sound levels and correct performance.

4.3 Calibration

4.3.1 Equalization

The equalization process consists on the design of a minimum-phase FIR filter to compensate the possible frequency deviations due to the directional microphone response and near-field effects due to its positioning close to the trumpet bell. The goal is to achieve a flat frequency response at the input of the convolution engine.

To this end, two trumpet players (one semi-professional and one amateur) were recorded in the anechoic room of the Erich Thienhaus Institute. They were asked to play an excerpt using the full frequency range of the instrument (up to the higher note they could achieve). The same microphone used in the virtual environment (Schoeps CCM 4V - see Fig. 4.12) was attached to the bell of the instrument. At the same time, an omnidirectional measurement microphone (NTi M2010) was placed approximately 2 meters away, aligned with the trumpet bell and its direction of maximum radiation, and used as a reference. The spectra of the recordings are compared against the reference microphone to obtain a compensation curve. Since the curves of the two musicians are slightly different, the average between those curves is computed and the resulting magnitude is used later to define the gain of the filter at different frequency bands.

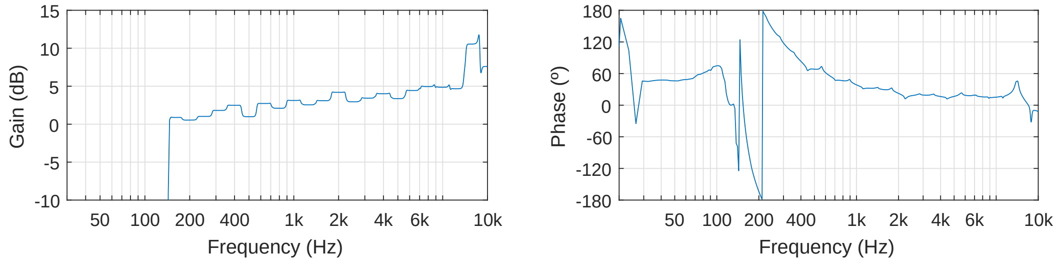


Fig. 4.14.: High-pass filter used to cut off noise from the valves of a trumpet.

The operation of the trumpet valves generates low frequency noise that could contribute to the creation of artifacts if it is not properly treated. To assess this issue, a recording of only valve noise is performed. When comparing the spectral envelopes of the previous recordings with the noise from valves it is found that the crossover frequency between the noise and the actual sound of the trumpet is around 160 Hz (see Fig. 4.13), which corresponds approximately to the frequency of the lowest note a trumpet can generate. Taking this into account, the filter gain below this frequency is modified to -60 dB, thus removing most of the energy contribution of the valve noise.

Finally, the FIR filter is generated using a parallel graphic equalizer [Ram+14] and saved as an audio file that will be loaded in the convolution engine. The magnitude and phase responses of the filter are shown in Fig. 4.14

4.3.2 Level calibration

When implementing a virtual environment where the live sound of a musician is mixed with electronically generated sound it is essential to ensure an appropriate relative level between the musician's sound and the room feedback. However, this is not a trivial task, and in many cases it is not possible to validate this calibration beyond the musicians' subjective experience.

To calibrate the sound level a musician was asked to perform a short excerpt which was recorded using the bell microphone and the sound pressure level of the sound was measured using an SPL meter placed near the right ear of the musician. Then the recorded sound was convolved with an impulse response containing the direct sound measured on stage with the previously described measurement setup. Then the convolved sound was played back into the room using the loudspeaker setup. The sound pressure level of the reproduced sound was measured and the gain of the system was modified until the level of the reproduced sound matched the real sound.

The calibration procedure relies on the similarity between the trumpet and the loudspeaker measurement properties, since the impulse responses used in the calibration procedure are measured with a directional source and imitating a musician setup on stage. The procedure was repeated with two different players achieving very similar results (< 1.5 dB of difference). In addition, before participating for the first time in an experiment, musicians were asked about the plausibility of the room loudness.

Fig. 4.15.: Measurement setup for the validation of the D3S auralizations.

4.4 Validation

To validate the implementation of the virtual environments and plausibility of the resynthesized aurizations, a set of measurements were conducted using the reproduction setup. The microphone array used in stage measurements was placed inside the listening environment and swept sine signals were convolved with the resynthesized SRIR of the measured rooms (see Fig. 4.15). Then, multichannel RIR were obtained that allowed the comparison of the frequency response functions of the real and auralized rooms, estimation of monaural room acoustic parameters and the analysis of spatio-temporal properties of the auralizations. The results are presented in Fig. 4.16 (FRF and room parameters), and in Fig. 4.17 (spatio-temporal response).

The frequency dependent error of the auralizations is comprised between $\pm$ 3 dB in the range of 200 Hz to 4 kHz. At lower frequencies the geometrical modes of the reproduction room affect to the reproduction response. At high frequencies (greater than 4 kHz), the early energy of the auralization is decreased approximately 6 dB with respect to the real room. This could be caused by non-coherent addition of wavefronts reproduced by triplets of sources in amplitude panning. The contrary happens for the late energy, which presents a systematic increase of high frequencies. This is known feature of auralizations based on SDM analysis. Since every single sample of the impulse response is represented as a Dirac delta in the resynthesis, this results in an increase of high frequencies [Ter+13]. An equalization procedure based on time-frequency analysis was introduced by Tervo *et al.* in [Ter+15a], improving the frequency response of the auralization considerably. However, this procedure was not available at the time of implementing the present environment and it is left for future work.

The reverberation time (T_{30}) and clarity (C_{80}) of the real and the auralized rooms present differences usually within one JND, except in specific cases. The T_{30} of the auralized rooms is considerably greater than the original one at high frequencies, due to the reason previously

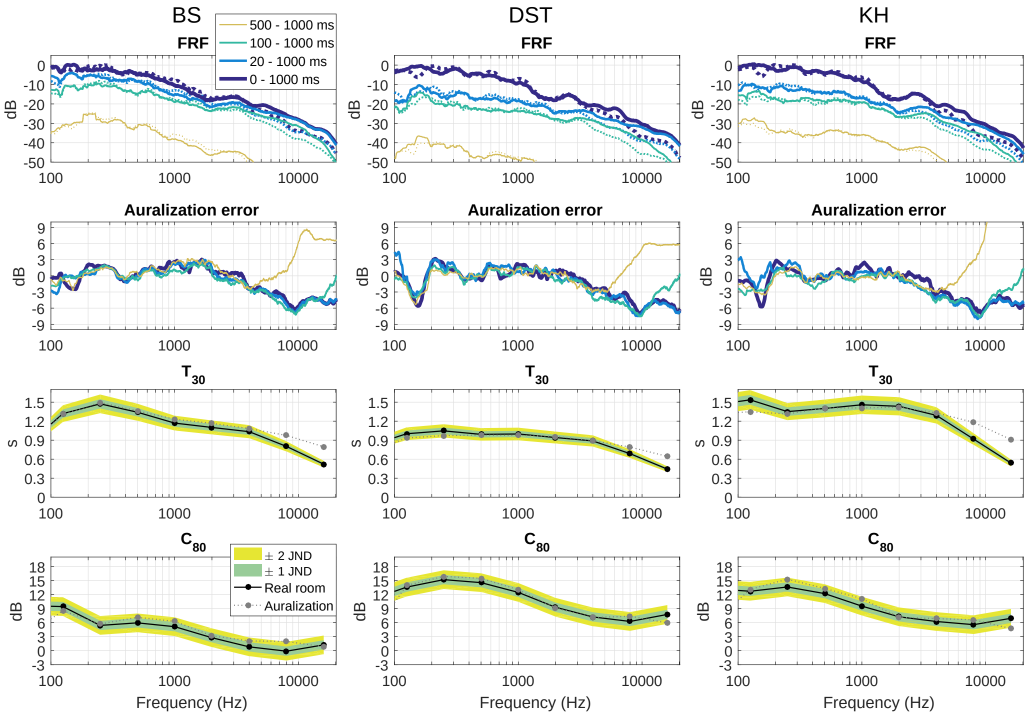


Fig. 4.16.: Frequency response, auralization error and room parameters of the original and auralized rooms. Solid and dashed lines on the first row correspond to the original and auralized rooms, respectively.

explained. The clarity of the auralized rooms is in all cases within 2 JND with respect to the original room.

The spatial distribution of early energy (between 10 and 80 ms) presents some discrepancies between the original and the resynthesized rooms. While the overall shape of the spatial representation is fairly similar, the analysis shows an increase of early energy. This is most noticeable in the case of the less energetic room - DST, which shows some prominent peaks at the bottom and back of the listening position. This is most probably caused by the non perfect anechoic conditions of the reproduction room, provided that its reverberation time at mid frequencies ranges from 50 to 90 ms. However, the spatial distribution of the late energy shows a good agreement between the original and the auralized room, presenting similar shape and energy levels. It should be expected that using a completely anechoic chamber for the reproduction would improve significantly the agreement between the original and the resynthesized rooms. In addition, the analysis error of the spatial analysis of the auralized rooms contain the analysis error of the real rooms plus the error of the auralized resynthesized rooms, thus leading to an increase of the analysis uncertainty.

4.4.1 Perceptual Validation

A group of semi-professional number players was invited to perform in the virtual environment and complete a survey as part of the experiments described in the next chapters. Detailed data regarding the musical background and the familiarization of these players with the auralized acoustics is included in Appendix C.

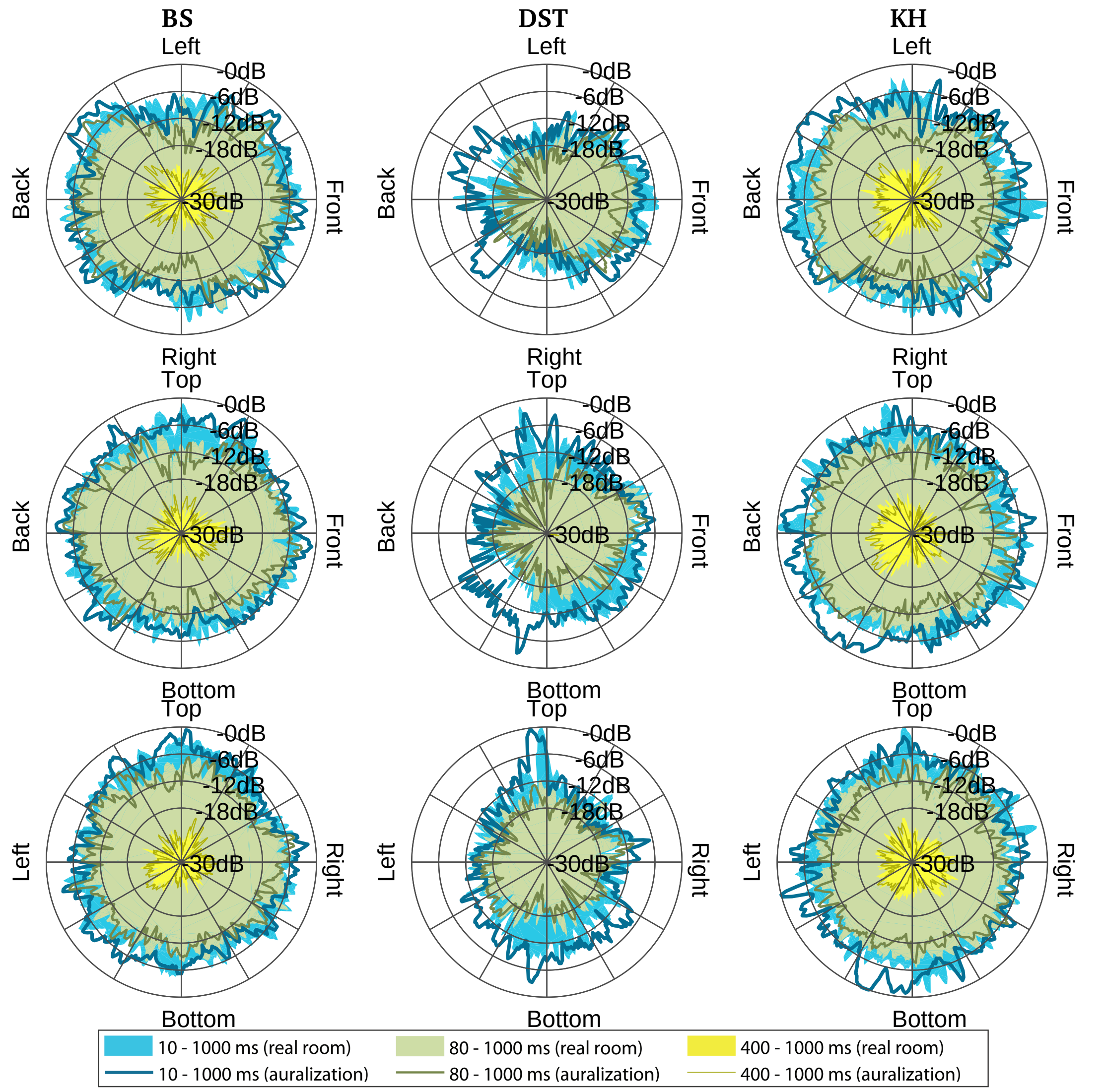


Fig. 4.17.: Spatio-temporal sound energy representations of the real rooms (filled regions) and the corresponding auralizations (solid lines).

At the time of the survey all the musicians except one had previously performed in the real rooms, and they were asked to rate the degree of realism of the resynthesized versions. Since the rating was an overall assessment of their impression, without the possibility of comparing the auralized and real rooms in real-time, the results should be interpreted as an assessment of degree of plausibility. The average score (on a scale of 100) was 81, with a standard deviation of 14, suggesting that the degree of realism (or plausibility) is sufficiently high to provide satisfactory acoustic feedback in real-time. In addition, the degree of intrusion of the setup (loudspeakers, seating arrangement, cables...) received an average score of 17, with a standard deviation of 25. Thus, it can be concluded that the degree of discomfort due to the experimental setup is rather small. In addition, all the musicians stated that they would be interested in practicing with such a system in a regular basis, as it contributes positively to their awareness of the surrounding acoustic conditions. The average degree of *usefulness* of the system in a context of musical training was 90, with a standard deviation of 16. The ratings of the players regarding the auralization system are collected in Tab. 4.5. A collection of comments made by the musicians is presented in Tab. 4.6.

Player	Auralization realism	*Usefulness* in musical training	Interested in regular use?	Intrusiveness of the setup
T1	70	100	Yes	70
T2	80	90	Yes	10
T3	90	100	Yes	10
T4	60	70	Yes	20
T5	70	90	Yes	0
T6	100	100	Yes	20
T7	100	100	Yes	0
T8	80	90	Yes	0
T9	70	100	Yes	0
T10	70	50	Yes	60
T11	100	100	Yes	0
Avg.	81	90		17
Std. Dev.	14	16		25

Tab. 4.5.: User ratings regarding the D3S auralization system.

Player	Comment
T9	The differences between real and auralized rooms are mostly physical (dimensions) and visual, rather than acoustically.
T9	Characteristics of sound are similar in real and auralized rooms.
T8	The sound of DST feels real, KH does not.
T11	The auralizations sound really realistic.

Tab. 4.6.: Players' comments regarding the auralization realism.

4.5 Further considerations

The presented system is able to provide plausible auralizations of measured rooms for trumpet players in real-time. By means of a directive source and a microphone array, an arrangement imitating a trumpet player on stage is used to capture a SRIR. Thus, if the same approach is considered for another instrument, the appropriate radiation characteristics and distances between source and receiver must be considered. Alternatives to obtain SRIR linked to specific musical instruments are the generation of directional sources using loudspeaker arrays [Pol15] or using artificial excitation of real instruments. However, this implies the conduction of acoustic measurements for every instrument, with its associated logistic complexity. Thus, an approach to generate SRIR for a number of instruments with different source configurations consists of using simulation methods [SK15].

In addition, given that the trumpet radiation is produced by a single aperture with axisymmetric shape, the miking approach is in this case rather simple. Nevertheless, most instruments have much more complex radiation characteristics, and at the moment most of the already implemented auralization systems use only one microphone. In order to capture the complete acoustic picture of the sound radiated by an instrument, approaches using multiple microphones should be considered.

5 Stage acoustics preferences of solo musicians

As concluded from a musicians' survey completed among trumpet players (see Tab. C.2), room acoustics are of great importance when performing both as an orchestra player (81/100) and as a soloist (75/100). However, one might expect that the acoustic requirements in an orchestral set-up and in a solo performance may differ greatly, as might do the acoustics required by a violinist and those required by a percussionist. The goal of this chapter is to present a pilot methodology to evaluate room acoustic conditions for solo players.

The virtual environment implemented during this project and presented in Chapter 4 is used to complete studies on stage acoustic preferences of solo trumpet players. The auralized rooms used in the investigation are the same previously presented - *BS*, *DST*, and *KH*, plus the quasi-anechoic conditions of the studio, *dry*. By having the possibility of switching between acoustic conditions in real-time it is possible to implement formal studies with musicians in a straightforward way, in controlled conditions.

To this end, two pilot experiments have been completed: "Performance Context" and "Directional Early Energy". In the experiment "Performance Context" the acoustical preferences of musicians under different performance contexts or situations is investigated, while in the "Directional Early Energy" test, the investigation focuses on the influence of the directions of early reflection on the perceived acoustic stage support.

5.1 Procedure and aparatus

Both tests consist of pairwise comparison tests, judging the subjective preference of all the acoustical conditions against each other in pairs. The order of presentation of pairs is randomized, and the test procedure is guided through a graphical user interface (GUI). The GUI, developed in Max/MSP and connected to the real-time convolution engine of the auralization system, presents the participant with the test question and choices, and performs background operations such as selection of appropriate SRIR for convolution and storage of test results.

The participants can switch in real time between the compared acoustic conditions and select the preferred choice using a MIDI controller. The GUI is presented to the participants in a monitor close to the playing position. An image of the experimental set-up from the musicians' perspective is displayed in Fig. 5.1. The main idea behind this experimental set-up is to allow the participants the complete the test autonomously, in order to avoid

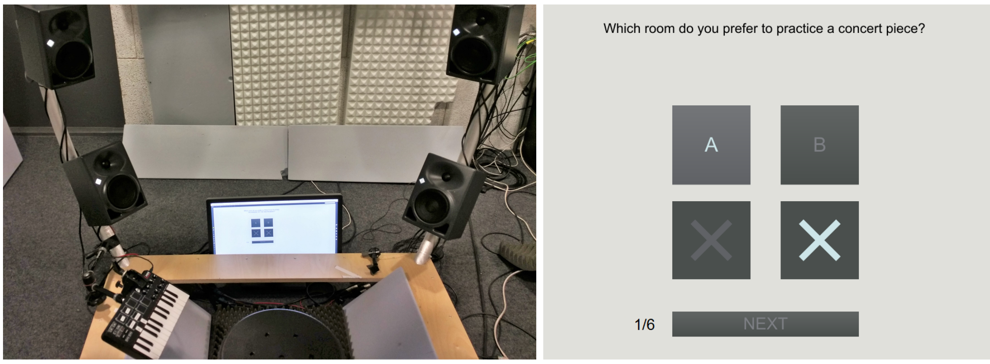


Fig. 5.1.: Aparatus (left) and screen capture of the GUI (right).

possible bias or influence of the experimenter on the results. The only interaction between the participants and the experimenter is limited at the beginning of the test, when the procedure is explained, or in case of a technical problem.

5.2 "Performance Context" test

A number of studies regarding stage acoustic preferences of solo musicians and ensembles are available [Gad89b; Jeo+14; Lim+13]. However, usually the vocabulary used in those studies refer to sound and acoustical qualities e.g. timbre, reverberance, or support - or generalistic terms, such as overall impression, or simply preference. One may intuitively conclude that the judged acoustical conditions refer always to the hypothetical situation of a musician performing in a concert. However, the amount of time dedicated to other activities such as instrumental practice or rehearsal is generally considerably greater than the actual concert performance.

The study "General Preference" aims at investigating the acoustical preferences of trumpet solo players under different performance contexts. To do that, musicians completed five consecutive pairwise comparison tests, using the same acoustic conditions (stimuli), and uniquely changing the test question. The following scenarios were tested:

- *Practice Technique*: In which room do you prefer to practice instrument technique?
- *Practice Concert*: Which room do you prefer to practice a concert piece?
- *Concert*: Which room do you prefer to perform in a concert?
- *Easiness*: In which room is it easier to perform?
- *Quality*: Which room does provide the best overall acoustic (or sound) quality?

The five different test scenarios were completed by 7 players in either one or two sessions, depending on the accumulated fatigue of the players. Every section lasted for approximately 10 minutes, and the judged pairs were compared twice with inverted orders (AB and BA) in

	Avg. dur. (s)	Median dur. (s)
Pract. Technique	79	42
Pract. Concert	45	35
Concert	33	28
Easiness	41	31
Quality	40	31

Tab. 5.1.: Duration of trials for the "Performance Context" experiment

order to test the consistency of the ratings. Every scenario consisted of 12 comparisons (6 pairs x 2 repetitions). While the presentation orders of the pairs was fully randomized, the order of the tested scenarios was the same in all cases (*Practice Technique*, *Practice Concert*, *Concert*, *Easiness*, and *Quality*). The measured average and median durations of every paired comparison are collected in Tab. 5.1. After the test, informal interviews were conducted in order to explore the subjective impressions of the musicians regarding the different acoustic conditions.

All the musicians participating in the experiments were students of the Detmold University of Music at the time of completion of the experiment. Given that all the participants in this test participated as well in the experiments described in Chapter 6, their musical background and personal data is summarized in Tab. 6.1 and the same abbreviation codes are used among the different studies to identify them.

5.2.1 Results

The results of the listening tests consist of a set of preference matrices, one for each test condition (see Tab. D.1 to D.5 in Appendix D). The Bradley-Terry-Luce (BTL) model [WS04] is applied on the matrices to estimate preference ratings of each room under each studied performance context. The result of the BTL analysis is shown in Fig. 5.2.

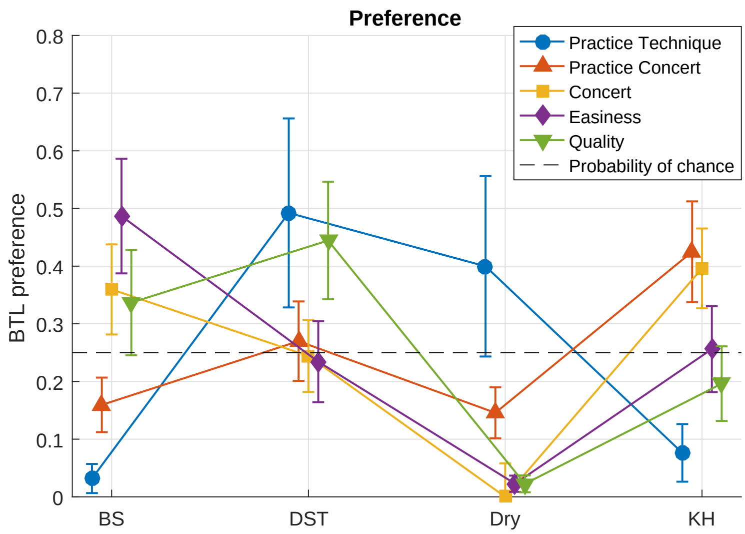


Fig. 5.2.: Estimated preference ratings (BTL) of stage acoustic conditions.

The results suggest that the estimated preference ratings differ significantly depending on the performance context. The rooms *DST* and *Dry* are significantly more preferred than *BS* or *KH* in the context of *Practice Technique*. The most preferred room for the condition *Practice Concert Piece* is *KH*, being the only one that presents an estimated preference significantly greater than the probability of chance. When judging acoustics regarding the performance of a *Concert*, rooms *KH* and *BS* present significantly greater preference ratings. The only room with a preference rating higher than the probability of chance regarding *Easiness* is *BS*. Finally, rooms *DST* and *BS* present higher perceived *Quality*.

In order to relate the preference ratings with the acoustical properties of the rooms, a second order polynomial regression model has been used to fit the average estimated preference values and the acoustic parameters of the room. As previous studies stated the importance of the reverberation time and the balance of early, late and total energy on the subjective judgments of room acoustics, the following parameters are used: RT_{20}, G_{all}, G_{early} and G_{late}. The use of G parameters, instead of ST is due to the use of a non-standardized measurement set-up, in terms of distance between source and receiver, and directionality of the measurement source. The preference values mapped as a function of the room parameters and the corresponding fittings with an adjusted R^2 greater than 0.6 are depicted in Fig. 5.3.

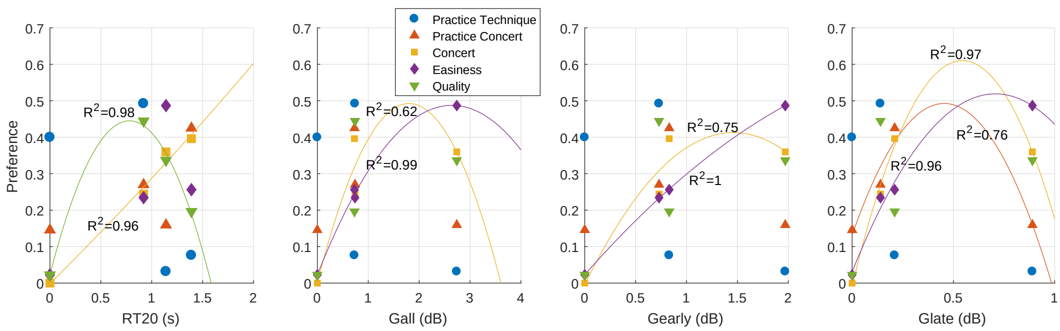


Fig. 5.3.: Second order polynomial fits of the room preferences and acoustic parameters.

Considering that only 4 fitting points are available from the experimental results, conclusions extracted from this analysis should be considered carefully. However, the results suggest the following:

- The preference of *Concert* conditions present a quasi-linear relationship with reverberation time (RT_{20}), suggesting that longer reverberation is beneficial for musicians performing in a concert. This goes in line with results obtained by Ueno *et al.* [UT03; Uen+05], who found out that reverberation time between 1.9 and 2.2 seconds was preferred over shorter reverberation times. In addition, Guthrie *et al.* confirm this trend in [Gut+13].

- The amount of early energy (G_{early}) presents a quasi-linear relationship with the *Easiness* of performance. In addition, higher values of total energy (G_{all}) and (G_{late}) present as well consistently higher values of preference, topping at around 2.7 and 0.9 dB, approximately. This suggests that especially higher early energy contributes positively to the comfort of the musician.

- *Quality* presents a quadratic relationship with RT_{20}. However, the perception of sound quality in a concert hall is a multi-dimensional [Lok14; KL17], at least from the listener point of view. Thus, one may consider that the rating of quality by musicians must respond as well to a combination of energetic, spatial and spectral aspects, and a parameter such as RT_{20} is insufficient to judge the overall sound quality.

5.2.2 Interviews

During the experiments the musicians could freely provide feedback about the acoustic conditions, and after the completion of the experiment informal interviews were completed. General comments provided by the musicians are summarized in Tab. 5.2. In addition, subjective information relative to every room is collected in Tables 5.3 to 5.6.

The subjective feedback provided by the musicians supports greatly the estimated preference results obtained by the formal test. When commenting about a specific room, several players explicitly reaffirm the kind of performance context that is benefited by those particular acoustics. A general comment shared by multiple players refers to the relationship between acoustic properties of the space and the ability to perform in a relaxed manner. In this sense, favorable acoustics lead to a state of relaxation that impacts positively on the musician psychological state and alleviates stress.

It is a much extended impression that a very dry environment (such as *dry* room) is generally beneficial to practice instrument technique, thanks to the high acoustic clarity of the room and the possibility of hearing small nuances related to articulation. However, the sound quality of the room is disliked by most of the players, and performing in such environment leads to a rapid increase of fatigue and is ultimately highly uncomfortable. For this reason, a room with high clarity and short reverberation, such as *DST* can retain the advantages of a very dry environment while helping to deal with the playing fatigue. In addition, *DST* is rated by some players as the best sounding room, or overall preferred.

According to the preference results, room *BS* is the best rated in terms of *easiness*, a result that is explicitly backed by the interview data provided by players T1 and T9, who state that this is the most comfortable room. In addition, players state that this room has some positive qualities, such as retaining some clarity and still having room, or being it a good room to play loud. However, it is not considered to be the best hall.

Room *KH* is overall preferred to perform concerts, and this is mentioned by the players. T1 considers that the sound is more projected into the hall, contributing positively to the phrasing. In addition, player T11 explicitly mentions the preference of this room for concert performances.

5.3 "Directional Early Energy" test

The methodology to manipulate the spatial characteristics of auralized impulse responses, presented in Sec. 4.2.1 is used to generate 5 variants of the resynthesized rooms:

"Performance Context" test	
Player	Comment
T2	A dry room is the best to study instrument technique, a bit more reverb to study a concert piece.
T3	It is easier to feel relaxed in a bigger room.
	An open room is preferred for slow pieces.
T4	It is easier to perform with comfortable acoustics.
	If the acoustics are good one enjoys more playing. Otherwise, it causes stress.
	There is no life without acoustics.
T5	Nice acoustics help on relaxing and enjoying the piece, also affecting the playing.
	Struggling to hear the performance properly causes nerves, increasing heart rate and being exposed to more mistakes.
T10	The musician prefers similar rooms for conditions concert practice and concert.
	The acoustics of a room can provided a wrong picture of one's performance.
T11	Appropriate acoustics allow one to be relaxed. Otherwise one focuses on correcting mistakes and not on making music.

Tab. 5.2.: Musicians' feedback regarding "General Preference" experiment.

Dry	
Player	Comment
T1	A dry room gives a feeling of oppression and leads to faster fatigue.
	The errors are easier to identify because when one stops blowing the sound immediately drops.
	This room is better to study. One can feel all the nuances of the articulation.
T2	Dislikes the room in general.
	It is good to practice technique, one can purge the bad habits.
T3	Likes the clarity of the dry sound.
	It's an interesting acoustic condition.
T4	A very dry room demands more resistance on the lips.
	Playing classical pieces is more demanding in this room, and it can be overcome by changing the articulation.
	Dry rooms are better to study technique passages, because the articulation can be easily heard.
	In this room one hears the true sound, acoustics tend to mask the actual performance.
T5	A trumpet player is more exposed in a dry room because the articulation is very clear.
T7	It is good to practice, but produces fatigue faster (specially on the lips).
	The sound is ugly.
T8	The sound is too dry - "tt sound".
T9	This is the most difficult to play in.
T10	It is strange to perform in this room when used to other acoustics.
	Some feedback is preferred, but this room is preferable over church-like acoustic to practice technique.
T11	It is easy to perceive imperfections and difficult to play nicely. One cannot play freely.

Tab. 5.3.: Musicians' comments regarding the room *Dry*

BS	
Player	Comment
T1	This is the most comfortable room to play in.
T7	Retains some clarity and has room, but it is not a great hall.
T8	BS is overall preferred over KH
T9	This is the most comfortable room to play in and one can hear more.
T11	The real BS is good when playing loud and hearing the room "vibrating", but not the best.
	It is a good room to practice.

Tab. 5.4.: Musicians' comments regarding the room *BS*

DST

Player	Comment
T2	This is the best room.
T5	This is the overall preferred room.
T7	This is the best sounding room.
	Jazz players are used to rooms like this.
T8	This is the best sounding room.
T9	The room is "perfect".

Tab. 5.5.: Musicians' comments regarding the room *DST*

KH

Player	Comment
T1	The sound bounces more and one can hear the performance better.
	The sound is more projected to the hall, but the reverberation is a bit too long.
	It is easier to do more phrasing.
T10	This room fits better for Ropartz, since it is a piece that depends on acoustics.
T11	This room is preferred for concerts.

Tab. 5.6.: Musicians' comments regarding the room *KH*

- *All*: Resynthesized SRIR without modifications.
- *Front-back*: Resynthesized SRIR with early reflections from front and back directions only.
- *Sides*: Resynthesized SRIR with early reflections from the sides only.
- *Front-back*: Resynthesized SRIR with early reflections from vertical directions only.
- *No-ER*: Resynthesized SRIR without early reflections

In all the auralized rooms the start of the late reverberation (unmodified energy) is set at 100 ms. The mixing time between early reflections and late reverberation is 45, 45, and 65 ms for the rooms *BS*, *DST*, and *KH*, respectively.

The goal of this experiment is to test whether the directional properties of early energy are relevant for the perception of stage support in solo musicians. The procedure of the experiment is based on pairwise comparisons of the different versions of each room, choosing which one provides better acoustic support. To ensure an adequate interpretation of the judged term, the musicians were instructed to choose the room that provided a better response in terms of hearing their own sound without difficulty or necessity of forcing the instrument, following the definition of Gade [Gad89b].

The test consisted of 30 comparisons (3 rooms x 10 pairs) fully randomized, and lasted approximately 30 minutes per participant. The experiment was completed by 5 participants, all of them trumpet students of the Detmold University of Music. The average and median measured durations of every trial are collected in Tab. 5.7.

	Avg. dur. (s)	Median dur. (s)
DST	38	28
BS	45	39
KH	40	27

Tab. 5.7.: Duration of trials for the directional early energy experiment

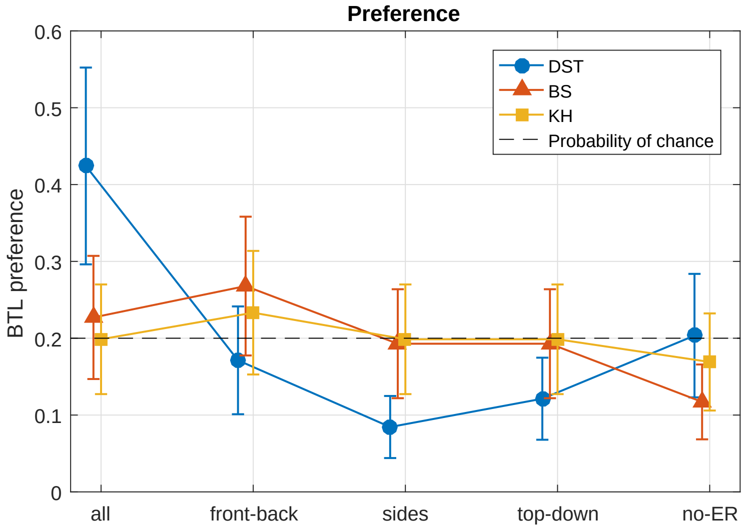


Fig. 5.4.: Estimated preference ratings (BTL) regarding the judgments of directional early support.

5.3.1 Results

The preference matrices obtained from the pairwise comparisons have been analyzed using a BTL model [WS04] and estimated preferences have been obtained. The results of the BTL analysis are displayed in Fig. 5.4.

The results of the preference ratings suggest that there are no statistical significant differences between the rooms with different directional early energy. Instead, the room *DST all* i.e. unmodified room, is significantly more preferred that its versions with reduced or modified early energy. Regarding the rooms *BS* and *KH*, one may argue that the version *front-back* presents a slightly higher preference rating, while *no-ER* is the least preferred. However, the results are not statistically significant and further testing with an increased number of subjects must be performed in order to verify those trends. Preliminary conclusions with the present results are that the directional properties of early energy are not relevant to judge the perceived stage support.

Analyzing the only significant results corresponding to room *DST* one possible explanation for the obtained preference ratings is that a minimal amount of early energy is required in order for the musician to experience appropriate acoustic support. Since *DST* is the least energetic room (smallest ST and G values among the compared rooms), the only variation of room that would fulfill the required amount of early energy is the original room (*all*), without any reduction. The average values of the estimated preference ratings are displayed as a function of the ST_{early} parameter in Fig. 5.5. There it can be observed that the most energetic (and preferred) version of *DST – all*, has lower amount of early energy than most of the *BS* variants. In addition, 4 out of 5 of the versions of *KH* have early energy values

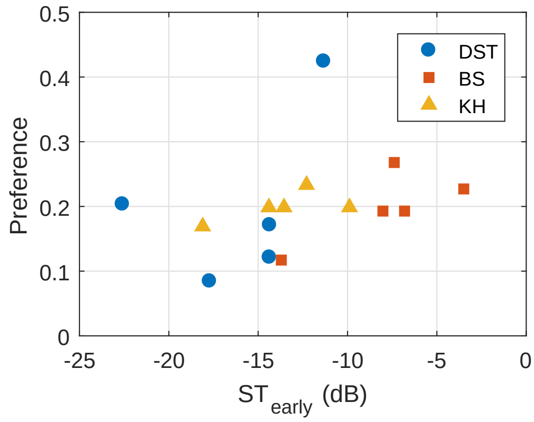


Fig. 5.5.: Estimated preference ratings (BTL) of perceived directional early support as a function ST_{early} parameter.

comprised between -15 and -10 dB, and very similar preference ratings. The same occurs for room *BS*, which has 4 of the versions with ST_{early} values comprised between -8 and -3 dB and preference values between 0.2 and 0.3. Considering that the just noticeable differences of ST_{early} are at the moment unknown, and the modified versions of room *DST* present generally lower ST_{early} values, a possible explanation is that the rooms *BS* and *KH* generally provide an appropriate amount of early energy, while the modified versions of *DST* do not, thus leading to a lower perceived stage support. However, it must be noted that this statement is in fact a hypothesis, and more tests should be performed in order to validate it.

5.3.2 Interviews

Although a formal interview process was not conducted in this experiment, the players were encouraged to feedback, and a discussion was started by some players initiative. In these cases, most of the comments were related to the difficulty to differentiate between many cases. Player T3 stated that they listened how the room "comes back" right after an articulated note, attempting to perceive how the sound between their instrument and the room was blended. In addition, player T10 stated that in most of the cases subtle differences in the early response of the rooms could be perceived, but generally they were not relevant for the playing conditions.

5.4 Discussion

The "Performance Context" test demonstrated that the performed activity or judged aspect is relevant to the study of preferred stage acoustics. Thus, it should be noted that the optimization of stage acoustics often responds to specific contexts, and the same acoustic properties that apply to a certain activities can be suboptimal in other cases. Nevertheless, the fact that higher levels of early energy contribute positively to the easiness of performance in solo players confirms Gade's findings [Gad89b], who stated that musicians generally prefer audible levels of early reflections when judging stage support. Since support is defined as the ability of hearing one's own sound without effort or forcing the instrument, *easiness* and perceived stage support could have a degree of comparability. In addition, the preference of longer reverberation times for *Concert* conditions, correlates well to the findings of Ueno [UT03] and Guthrie [Gut+13].

However, it is worth noting that studies found in literature often rate aspects such as *Overall Acoustic Impression*, *Quality* or simply *Preference*, terms that can have individual interpretations and lead to contradictory results. In addition, during both experiments, the character of the musical pieces interpreted during the exploration of acoustics varied depending on the test. For example, while completing the test *Practice Technique* musicians generally played scales, *études* with clear articulation or passages with high technical complexity. Contrarily, while testing *Practice Concert* or *Concert*, they tended to play more lyrical and expressive passages. Interestingly, in some cases the technical complex passages used in *Practice Technique* and lyrical passages from the later tests could be different sections or movements of the same pieces. Finally, most of the players tended to play passages with fast and clear articulation during the experiment "Directional Early Energy", in order to explore the early response of the room.

The average duration of every trial in both tests are collected in Tab. 5.1 and 5.7. An ANOVA analysis revealed that the trial duration of the very first test in the "Performance Context" experiment (*Practice Technique*) was statistically significantly higher than the rest of the trials. A possible explanation is that the musicians experience a familiarization process with the virtual acoustics and test procedure, and after the first test the comparison task becomes easier. The average duration of every trial in the tests "Performance Context" and "Early Directional Energy" are comparable, which suggests that although not significant differences are found between stimuli in the latter, the task complexity is similar.

Although some inferences have been made, the comparability of results with past studies that used virtual acoustic environments is subject to some limitations. The main constraint is that the estimation of room acoustic parameters in virtual environments is subject to high uncertainty for many reasons: the use of non-standard sources with non-omnidirectional radiation properties affect substantially the estimation of room parameters [Lar+16], the use of non-standard source to receiver relative distances affects the direct to reverberant ratios in energy measurements, and the fact that instruments must be miked in virtual spaces means that the auralization method does not consider energy radiated in directions not covered by the microphones. Given those conditions, it is necessary to establish new methodologies to measure and compare the room acoustical properties in virtual spaces, and generally the measurement process description is not detailed enough to allow reproducibility of acoustic measurements. These aspects, together with the calibration issues already present and discussed in the measurement of parameters such as ST [Wen+12; Dam11] pose a limitation in the comparability of studies regarding the quantification of energetic aspects of rooms. However, considering the logistic and experimental advantages of laboratory experiments over *in-situ* studies, adequate solutions to these issues, together with a correspondent update on the standard ISO 3382 [ISO09] would suppose a benefit in this research field, by allowing the implementation and comparison of similar experiments in different virtual acoustic systems.

6 Performance adjustments due to room acoustics

Room acoustics constantly modify the sound generated by a musician's instrument, thus impacting directly on the performance. The goal of this chapter is to determine whether musicians – either consciously or unconsciously – modify their performance depending on the room acoustic conditions. To this end, two experiments have been implemented: the first one features trumpet players performing in the D3S, while in the second experiment organ players are recorded using a MIDI interface in the Detmold Konzerthaus, which has a RAES installed.

6.1 Trumpet performance in virtual environments

The study of trumpet performance has been carried out through the completion of playing experiments with trumpet players in the virtual acoustic environment presented in Chapter 4. This section presents the utilized experimental procedure, analysis of performance data and subjective feedback from musicians.

6.1.1 Experiment description

The experiments were conducted in the WFS Studio of the Erich Thienhaus Institute and consisted of several recording sessions with musicians. The main task was to perform a music excerpt under different acoustic conditions. The recordings were then recorded to extract and analyze musical features for further analysis.

Eleven trumpet players participated in the experiments (see Tab. 6.1 and C). All the musicians were students at the Detmold University of Music at the time of the experiments, and their level ranged from third semester of bachelor studies to first semester of master. The musicians were asked to prepare at least two excerpts with different musical characteristics. They

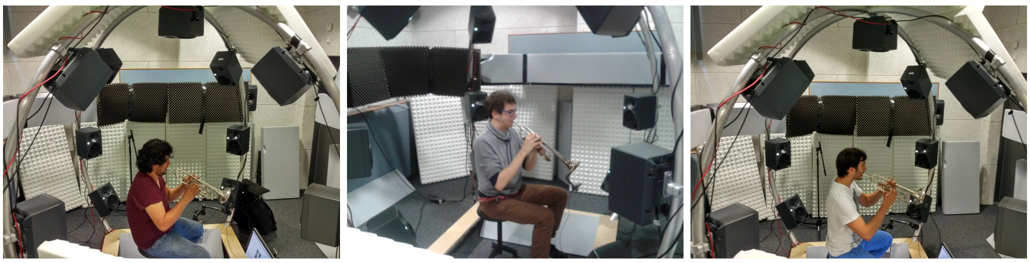


Fig. 6.1.: Trumpet players during the experiment.

Player	Level	No. Sessions	No. Recordings
T1	Bachelor (finished)	1	16
T2	Bachelor (5th semester)	3	55
T3	Bachelor (3rd semester)	3	69
T4	Master (1st semester)	2	32
T5	Bachelor (finished)	1	25
T6	Bachelor (7th semester)	1	11
T7	Bachelor (7th semester)	3	73
T8	Bachelor (5th semester)	1	17
T9	Bachelor (7th semester)	1	20
T10	Bachelor (4th semester)	2	30
T11	Bachelor (5th semester)	1	16

Tab. 6.1.: Trumpet players participating in the experiments.

were free to choose the pieces, with the only requirement of being able to perform the piece without technical or interpretative difficulties, in order to avoid learning effects during the experiment.

At the beginning of the session the musician is instructed to seat at the center of the loudspeaker set-up, and a microphone is installed on the trumpet bell. Afterwards, a sound check is completed, ensuring the appropriate calibration values and plausibility of the auralized acoustics. Once the system is configured, the musician is introduced to the experiment, receiving basic information about the auralization system and the goal of the experiment. Then, together with the present researcher, the musical excerpts to be performed are selected and the recording session starts.

During the recording session the musician performs the same excerpt under different acoustic conditions, and the performances are recorded. The acoustic conditions are selected randomly, and musicians are allowed unlimited time to become familiar with a new room before starting the actual recording. This process is repeated for every piece, and the number of repetitions and duration of the session is adapted depending on the fatigue of the participant.

After completing the recordings, musicians are asked to fill a survey providing information related to their musical experience, as well as feedback regarding the experiment and realism of the acoustic conditions. Finally, an interview is conducted, asking musicians about the implemented performance adaptations under different acoustic conditions. The interviews are a combination of closed questions with cooperative conversation. The closed questions explicitly refer to the modification of the following musical interpretative aspects: *tempo*, articulation, dynamics, and expressivity. The initial responses of the musicians are then further discussed through cooperative conversation. Interviews were conducted in English, German, and Spanish, and the qualitative information was then translated and summarized. The results of the interviews are presented together with the signal analysis of the recorded performances in Tab. 6.6 to Tab. 6.16.

Four acoustical conditions were used during the experiments: *BS*, *DST*, *KH* and *dry*. The first three rooms and their auralization characteristics have already been presented in Section 4.1.2. The *dry* condition corresponds to the quasi-anechoic acoustics of the WFS Studio, with the auralization system off. The room acoustic parameters of the experimental

Room	EDT	T_{20}	T_{30}	C_{80}	ST_{early}	ST_{late}	G_{early}	G_{late}	G_{all}	TS
Dry	0.01	0.05	0.07	62.76	-29.23	-55.43	0.03	0.00	0.03	0.00
BS	1.22	1.21	1.25	4.01	-3.06	-3.78	1.84	1.63	2.97	0.06
DST	0.52	0.96	0.97	11.63	-9.34	-12.15	0.70	0.37	0.99	0.02
KH	1.10	1.30	1.41	9.76	-9.03	-10.16	0.61	0.49	1.03	0.03

Tab. 6.2.: Room acoustic parameter of the experimental conditions, averaged over octave bands from 250 Hz to 4 kHz.

conditions are presented in Tab. 6.2 averaged over octave bands ranging from 250 Hz to 4 kHz. The Pearson correlation coefficients between the computed acoustic parameters are presented in Tab. 6.3. Given the high correlation between the computed room parameters, only EDT, T_{30} and G_{all} will be used in further analysis, provided that they can be computed with lower uncertainty than others e.g. ST_{early} [Dam09], thus allowing comparison of results in potential further studies.

	EDT	T_{20}	T30	C_{80}	STearly	ST_{late}	Gearly	G_{late}	Gall	T_S
EDT	1.00									
T_{20}	0.94	1.00								
T_{30}	0.94	1.00	1.00							
C_{80}	-0.88	-0.97	-0.96	1.00						
ST_{early}	0.9	0.96	0.94	-0.99	1.00					
ST_{late}	0.89	0.97	0.96	-1.00	0.99	1.00				
G_{early}	0.8	0.69	0.66	-0.75	0.84	0.77	1.00			
G_{late}	0.8	0.64	0.62	-0.68	0.78	0.71	0.99	1.00		
G_{all}	0.82	0.69	0.67	-0.74	0.83	0.77	1.00	0.99	1.00	
T_S	0.88	0.75	0.73	-0.77	0.85	0.79	0.98	0.99	0.99	1.00

Tab. 6.3.: Correlation coefficient between the computed room acoustic parameters.

6.1.2 Performance Dimensions

The 44 low level audio features presented in Sec. 3.2.2 are extracted from all the recordings completed during the experimental sessions. As opposed to MIDI analysis, where the performance is in fact coded and available in a data stream, when analyzing audio signals it is *a priori* not known which audio features are representative of the performance changes implemented by musicians. In addition, it is expected that several features are strongly correlated. Thus, in order to analyze which audio features are indeed relevant and reduce the dimensionality of the dataset, a Dual Multiple Factor Analysis (DMFA) is performed, using the package FactoMineR for R [Lê+08].

The reason for the implementation of DMFA, instead of a traditional Multiple Factor Analysis (MFA) lies within the nature of the dataset. While MFA is designed to reduce the dimensionality of a dataset composed by several variables collected on the same number of observations, the present data is composed by multiple observations (audio recordings) measured on the same variables (audio features). In this case, a matrix **X** of dimensions N (number of recordings) by F (number of features) is the result of the concatenation of K sets of observation matrices $\mathbf{X}_k$. An observation matrix is defined as all the recordings of the same piece (including all acoustic conditions) effectuated by one player in the same day.

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{[\mathbf{1}]} \\ \vdots \\ \mathbf{X}_{[\mathbf{k}]} \\ \vdots \\ \mathbf{X}_{[\mathbf{K}]} \end{bmatrix} \tag{6.1}$$

The grouping results in 46 groups organized in a hierarchical structure with three levels: player, session, and musical piece. The structure is presented in Fig. 6.2. This organization is based on the assumption that every group is understood as an independent set of observations. However, it is unknown whether the observations of a player performing the same piece in different days are independent. Nevertheless, since a PCA is performed on each group, as a first step of the MFA, this becomes irrelevant, and simplifies the issue of data centering. Given that differences in means between tables have an effect on the analysis, the centering of the variables must be considered in DMFA [Abd+13]. With the presented hierarchical structure, the dataset contains differences which are only meaningful within the same table e.g. the same player performing the same piece in different room, during the same session. Hence, the data is centered within each table individually. Finally, the variance of all 44 low level audio features is scaled to one during the MFA, ensuring an equal weight of every extracted feature in the generation of the resulting dimensions.

The MFA results in 44 dimensions with an associated explained variance. The first 10 dimensions account for approximately the 80% of the explained variance (see Fig. 6.3). The contributions of each extracted audio feature to the MFA dimensions is displayed in Tab. 6.4. As extracted from this data, the first four dimensions, which account for approximately 58% of the explained variance, present strong relationships with some of the analyzed audio features.

Dimension 1 is related to energetic and spectral features, and thus related to the overall level of the performance and the timbral aspects of the generated sound. In this sense, the higher values in Dim. 1 indicate a louder performance with overall brighter sound. The energetic features that contribute most to the dimension (in order of importance) are: *LUFSlinear*, *rmsA*, *rms*, *rmstoneenv*, *SFvar*, *SFmean*, and *rmsphraseenv*. On the spectral side, the most important contributors are: *spectcentroid*, *spectbright2000*, *MFCC1*, *spectbright*, *spectrolloff*, *spectskewness*, *spectentropy*, *spectspread*, *MFCC2*, and *spectflatness*.

Dimension 2 is composed of features related to the temporal-energy characteristics of the sound, resulting from the analysis of the amplitude envelope of the signals. The main contributors to the dimension are: *lowenASR*, *enventropy*, *LUFSstd*, *LRAlinear*, and *lowener*. The nature of the contributing features suggests that this dimension is related to dynamic variations of the performance. However, these features are mostly related to the range of the dynamics, and do not provide information about the temporal structure of those dynamic

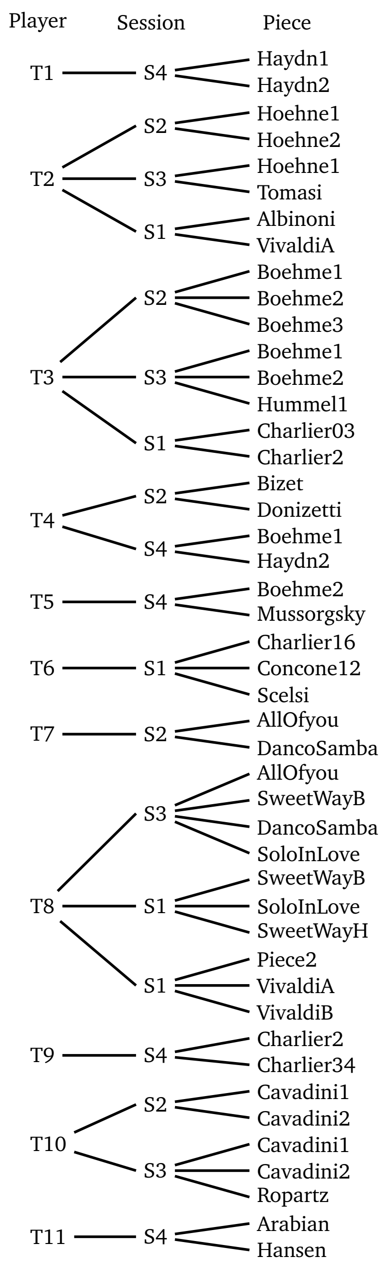


Fig. 6.2.: Hierarchical structure of the trumpet recordings dataset.

	Dimension							
Audio feature	1	2	3	4	5	6	7	8
'rms'	**0.89**	-0.24	-0.05	0.05	0.26	-0.19	-0.01	0.06
'rmsA'	**0.91**	-0.23	-0.04	0.05	0.22	-0.19	0.00	0.01
'LUFSlinear'	**0.92**	-0.18	-0.03	0.06	0.24	-0.19	-0.02	0.02
'LUFSstd'	0.00	**0.81**	0.07	-0.01	0.14	-0.13	0.18	-0.10
'SFmean'	**0.85**	-0.30	-0.17	-0.03	0.13	-0.09	-0.04	0.00
'SFvar'	**0.88**	-0.11	-0.06	0.02	0.24	-0.12	-0.05	0.06
'LRAlinear'	0.01	<u>0.74</u>	0.14	0.10	0.25	-0.13	0.14	0.07
'peak2rms'	0.21	0.47	0.17	-0.07	-0.20	0.20	0.15	0.06
'envspread'	-0.19	-0.26	<u>0.62</u>	0.48	-0.01	-0.09	-0.04	0.10
'envflatness'	-0.22	-0.51	-0.14	-0.20	0.03	0.23	0.37	0.20
'envcentroid'	-0.19	-0.10	<u>0.64</u>	0.46	0.02	-0.02	0.02	-0.01
'envskewness'	0.02	0.40	0.02	-0.09	0.33	-0.15	0.04	0.08
'enventropy'	-0.14	**-0.82**	-0.11	-0.10	0.02	0.21	0.16	0.17
'rmstoneenv'	**0.89**	-0.24	-0.05	0.05	0.26	-0.19	-0.01	0.06
'toneenvar'	0.59	0.44	0.07	0.15	0.45	-0.26	0.05	0.10
'rmsphraseenv'	**0.80**	-0.26	-0.01	0.09	0.35	-0.20	0.04	0.13
'lowener'	-0.08	<u>0.69</u>	0.01	-0.13	0.07	-0.04	0.23	0.04
'lowenASR'	0.00	**0.83**	0.06	-0.01	-0.01	-0.21	0.11	-0.16
'spectcentroid'	**0.93**	0.07	0.08	0.03	-0.22	0.15	0.04	-0.15
'spectbright'	**0.91**	-0.01	0.02	0.09	-0.14	0.12	0.05	-0.25
'spectbright2000'	**0.93**	0.01	0.05	0.05	-0.18	0.16	0.06	-0.17
'spectrolloff'	**0.90**	0.09	0.08	0.01	-0.20	0.16	0.08	-0.10
'spectskewness'	**-0.90**	-0.05	-0.06	-0.02	0.20	-0.09	-0.10	0.12
'spectspread'	**0.75**	0.30	0.16	-0.08	-0.27	0.19	-0.10	0.24
'spectflatness'	<u>0.60</u>	0.30	0.16	-0.06	-0.18	0.23	-0.20	0.34
'spectentropy'	**0.82**	-0.03	0.06	0.03	-0.17	0.13	0.04	-0.11
'MFCC1'	**-0.93**	-0.07	-0.09	-0.05	0.18	-0.17	-0.02	0.15
'MFCC2'	<u>-0.67</u>	-0.33	-0.15	0.10	0.38	-0.03	-0.04	-0.21
'MFCC3'	0.03	-0.01	-0.11	-0.06	0.18	0.23	0.51	-0.05
'MFCC4'	0.27	-0.18	-0.02	0.11	0.32	0.32	-0.02	-0.10
'MFCC5'	0.05	0.12	-0.05	0.11	0.39	0.53	0.17	0.14
'MFCC6'	-0.11	0.04	-0.03	0.24	0.49	0.34	0.05	-0.38
'MFCC7'	0.18	0.32	0.06	0.07	0.21	0.43	-0.21	0.50
'MFCC8'	-0.26	0.09	-0.05	0.11	0.31	0.47	-0.16	-0.19
'MFCC9'	0.18	0.15	-0.03	0.10	0.22	0.18	-0.47	0.04
'ZC'	0.42	-0.42	-0.06	-0.11	-0.13	-0.19	0.13	0.04
'length'	-0.20	-0.18	<u>0.72</u>	0.55	0.01	-0.05	0.03	0.04
'tempomedian'	0.05	0.19	<u>-0.62</u>	-0.41	0.03	-0.03	-0.17	-0.07
'tempomean'	0.17	0.15	**-0.87**	-0.11	0.01	0.02	-0.04	0.05
'tempospread'	-0.03	0.11	<u>-0.64</u>	<u>0.71</u>	-0.17	-0.02	-0.04	0.03
'tempoflatness'	0.07	-0.11	0.54	**-0.78**	0.18	0.03	0.00	-0.03
'temposkewness'	-0.02	-0.02	-0.18	0.26	-0.07	0.02	0.52	0.16
'tempokurtosis'	-0.01	0.03	-0.15	0.02	0.01	-0.09	0.14	0.41
'tempoentropy'	0.05	-0.10	0.56	**-0.78**	0.19	0.02	0.02	-0.03

Tab. 6.4.: Contribution of the audio features to the construction of MFA dimensions. Absolute values higher than 0.75 and 0.60 are shown in bold and underlined font, respectively.

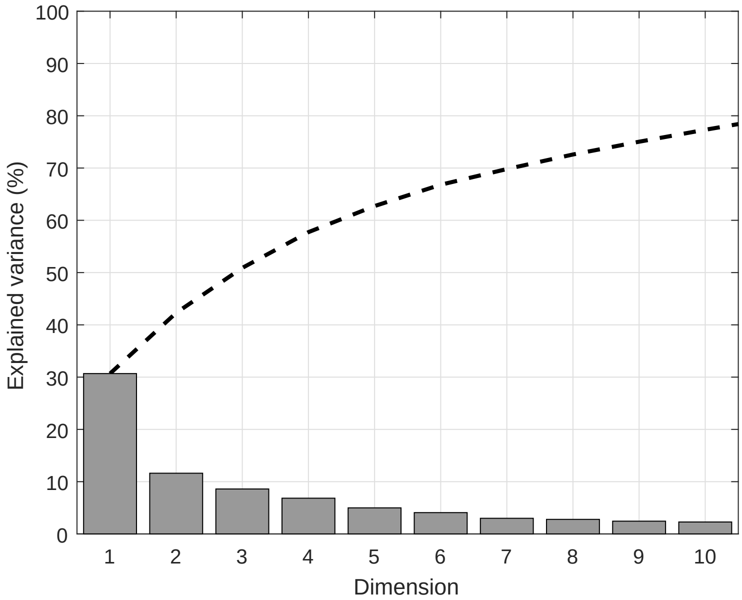


Fig. 6.3.: Explained variance of the MFA dimensions of trumpet performance.

variations. Overall, performances reporting higher values in Dim. 2 present overall a higher dynamic range than those with smaller values.

Dimension 3 relates to overall duration and *tempo* of a performance. The higher contributor to this dimension is *tempomean*. Other contributors are *length*, *envcentroid*, *tempospread*, *envspread*, and *tempomedian*. Given that the contribution of *tempomean* to Dim. 3 has negative sign, this indicates that higher values correspond to slower performances.

Dimension 4 has three features that contribute most to it: *tempoentropy*, *tempoflatness* and *tempospread*. These characteristics are statistical features extracted from the *tempo* curve of a performance. Given that *tempospread* contributes positively, while *tempoentropy* and *tempoflatness* present a negative contribution, suggests that higher values of Dim. 4 correspond to a performance with more *tempo* variations than those with smaller values. However, and analogously to Dim. 2, the fine nature of those variations can not be analyzed from the values of the dimension.

The rest of the dimensions do not present clear contributors, and they rather appear to be an heterogeneous mixture of several audio features.

6.1.3 General trends

A correlation analysis has been performed relating the MFA scores of all the recorded performances and the room acoustic parameters of the rooms used in the experiments. The values are reported in Tab. 6.5.

A moderate significant correlation (approximately 0.33) has been found between all the room parameters and the first performance dimension (overall level and timbre). The fact that all room parameters show greatly similar correlation values with the performance dimension, suggests that musicians tend to compensate for the reverberance of the room. In this sense, more energetic rooms, or rooms with longer reverberation lead to a decrease of the playing level, inducing as well a darker timbre.

Weak significant correlations (approximately 0.11) are found between energy parameters (G_{early}, G_{late}, G_{all}) and the *tempo* dimension. This suggests that more energetic rooms incite players to slightly decrease the tempo of the performance.

Finally, a weak correlation (approximately 0.24) is found between all room acoustic parameters and dimension 7. However, dimension 7 is composed of many audio features, and is not directly related with any of the typical performance aspects and it remains unknown what is the actual meaning of this correlation on the performance adjustments.

	Dimension							
Parameter	1	2	3	4	5	6	7	8
EDT	-0.34**	0.06	0.08	-0.01	0.01	0.01	0.24**	0.07
T_{20}	-0.32**	0.08	0.05	-0.01	0.04	0.03	0.24**	0.04
T_{30}	-0.31**	0.08	0.05	-0.01	0.04	0.03	0.24**	0.04
G_{early}	-0.34**	0.01	0.11*	-0.05	-0.04	-0.02	0.22**	0.07
G_{late}	-0.34**	0.01	0.12*	-0.04	-0.05	-0.02	0.22**	0.08
G_{all}	-0.35**	0.01	0.11*	-0.04	-0.04	-0.02	0.23**	0.07
ST_{early}	-0.34**	0.06	0.07	-0.03	0.01	0.02	0.25**	0.04
ST_{late}	-0.33**	0.07	0.06	-0.03	0.02	0.03	0.25**	0.04
T_S	-0.35**	0.02	0.11*	-0.04	-0.03	-0.02	0.23**	0.08
C_{80}	0.33**	-0.07	-0.06	0.02	-0.03	-0.03	-0.24**	-0.03

Tab. 6.5.: Correlation coefficient between measured room parameters and performance dimensions generated by MFA (* $p<0.05$; ** $p<0.01$).

The MFA scores corresponding to each of the recorded performances have been mapped on the first 4 dimensions of the MFA space (~58% of explained variance). Additionally, 95% confidence ellipses are included (see Fig. 6.4). The graphs show that the confidence ellipses of rooms *DST* and *KH* present a great overlap on dimensions 1 and 2, while rooms *dry* and *BS* present greater differences both in mean and area of the ellipses. However, the differences in mean are mostly present in dimension 1, suggesting that the performances recorded in rooms *DST* and *KH* are more similar in terms of sound level than the ones recorded in *dry* or *BS*, that present the highest and lowest sound level, respectively. Regarding dimensions 3 and 4, all the ellipses are highly overlapped, suggesting that there are not significant differences in terms of overall *tempo* or *tempo* variations generalized among all musicians.

Partial Clouds

In order to compare the behavior of each player individually in relation to the general trends, partial clouds are generated projecting the partial factor scores of every player into the MFA dimensions (see Fig. 6.5 and 6.6). The room acoustic conditions corresponding to every performance are used in the MFA as a supplementary table, and mapped on the resulting dimensions as centroids of every partial cloud.

The rooms that present higher differences in Dimension 1 (overall performance level) are *dry* and *BS*. All the partial variables (individual players) present lower values for *BS* than *dry*,

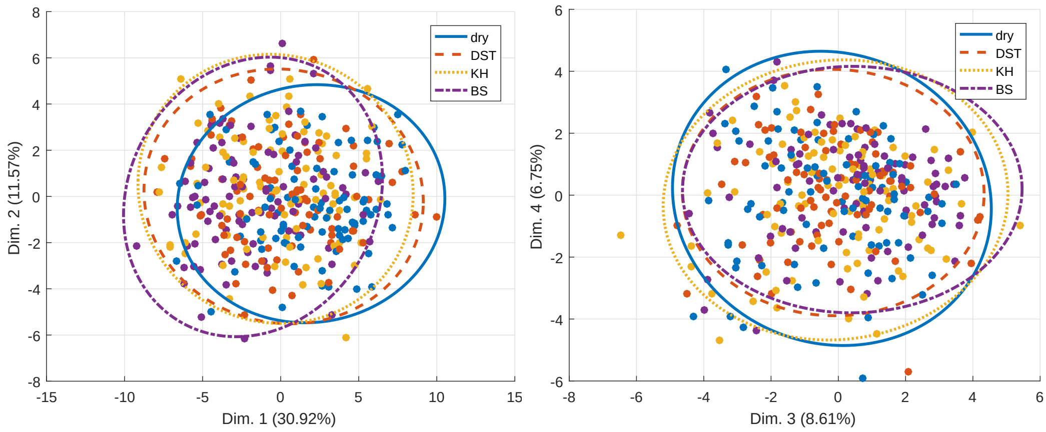


Fig. 6.4.: Individual recordings and confidence ellipses (95% interval).

with the exception of player T4. At the same time, rooms *KH* and *DST* are both fairly close and located around the origin of coordinates. Considering that the total energy of the rooms (G_{all}) is most different for rooms *BS* and *Dry*, while *DST* and *KH* present similar values, this suggests that the general tendency of the musicians is to reduce the level of the produced sound according to the increase of overall energy of the room.

Regarding dynamic variations (Dimension 2), the higher differences are between rooms *KH* and *dry*. These rooms happen to be the ones with highest differences in reverberation time (T_{30}). All the players except T6 present higher values of Dim. 2 in recordings performed in *KH* than in those performed in *dry*. However, in this case the differences are much smaller compared to those of Dim. 1, and a more thorough analysis is needed to determine whether they are significant or not. Centroids of rooms *BS* and *DST* show similar values, and overall, the deviation of individual players from the centroids is considerably large, resulting in non significant differences between performances in different rooms.

The individual behaviors regarding dimensions 3 and 4 (overall *tempo*, and *tempo* deviations, respectively) seem to be largely individual and the analysis from the partial cloud representations is not intuitive. However, although not statistically significant, the overall *tempo* (Dim. 3) of performances executed in *BS* seems to be slightly slower than those recorded in the rest of the rooms.

6.1.4 Individual players

This section presents the performance adjustments of every player under different acoustic conditions in terms of correlation between performance dimensions and room acoustic parameters, as well as partial clouds mapping every recorded piece on the reduced MFA space. The information is presented in Fig. 6.7 to 6.17. The correlation results include 95% confidence intervals. In this sense, correlation values with confidence intervals that do not cross the horizontal axis refer to statistically significant correlations at 5% level ($p<0.05$).

Additionally, the summarized qualitative data extracted from the personal interviews is presented separately for each musician in Tab. 6.6 to Tab. 6.16.

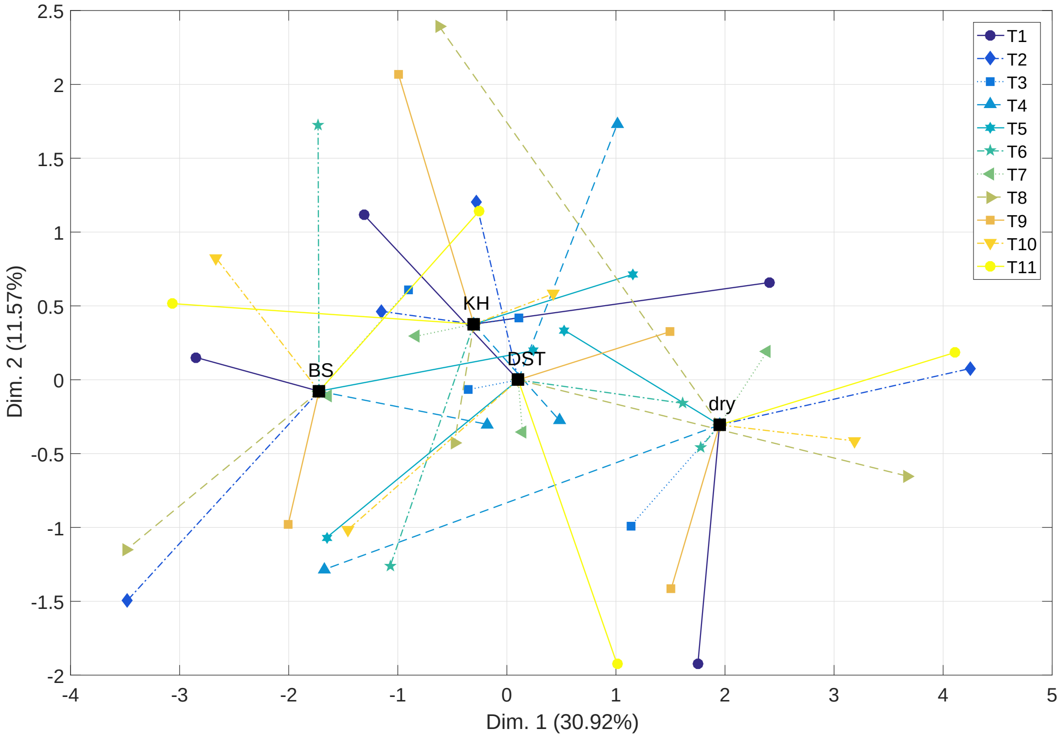


Fig. 6.5.: Partial factor scores projected into the MFA dimensions 1 and 2.

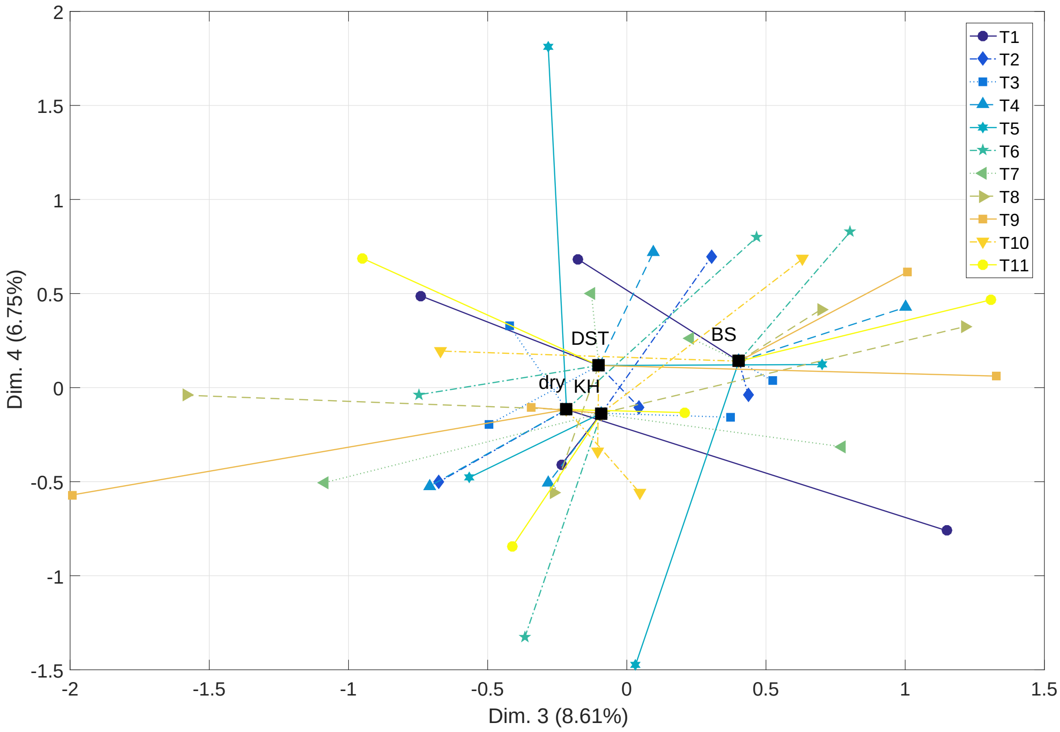
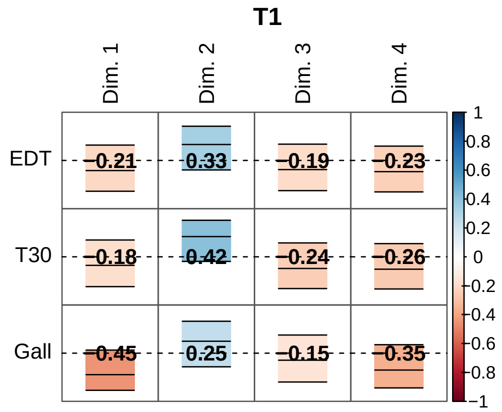


Fig. 6.6.: Partial factor scores projected into the MFA dimensions 3 and 4.

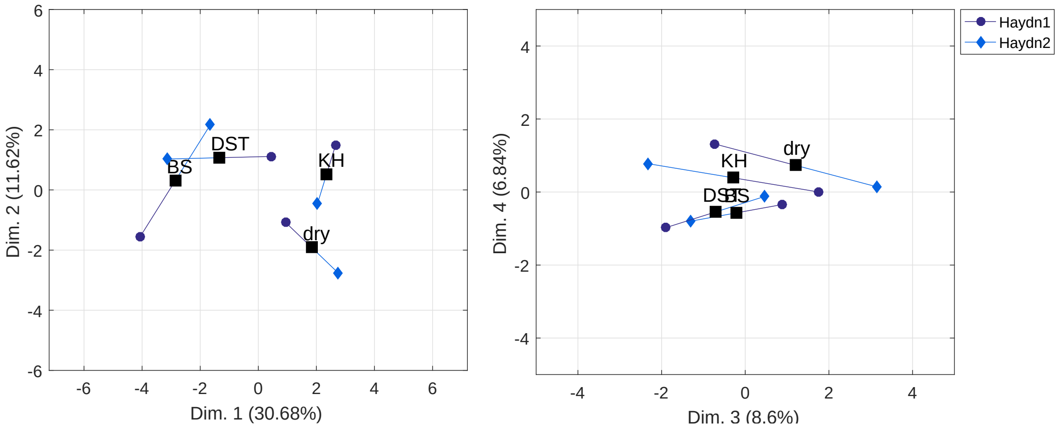
Player T1

Player T1 (see Fig. 6.7) does not present significant differences in performance behavior when performing under different acoustic conditions. Overall *tempo* and *tempo* variations (dim. 3 and 4) do not seem to follow a clear trend, while the overall performance level (dim. 1) and dynamic variations (dim. 2) present correlations close to significance with regard to G_{all} and T_{30} parameters (-0.45 and 0.43, respectively).

The partial cloud show that performances in rooms *dry* and *KH* tend to present higher overall level (dim. 1), and the centroid of performances recorded in room *dry* present slightly lower dynamic variation (dim. 2) levels. Contrarily to the results of the performance analysis, the player mentions that dynamics must be exaggerated in drier rooms in order to achieve the same musical idea as in more favorable acoustic conditions (see Tab. 6.6).



(a)



(b)

Fig. 6.7.: Results of player T1

Performance aspect	Adjustments
Tempo	More attention must be payed to tempo in a dryer room.
Dynamics	Need to exaggerate dynamics in a dry room to achieve the same musical level as in another room. The overall level does not change, but the idea does.
Articulation	The balance between air flow and articulation is harder in dry rooms. A wrong balance between tongue movements and air flow causes noticeable artifacts (described as "clac-clac").
Expressivity	One must put more effort in a dry room to achieve the same expressiveness of a longer reverberation and nicer room. In an appropriate room, one is more free to achieve a good phrasing.

Tab. 6.6.: Interview responses of player T1

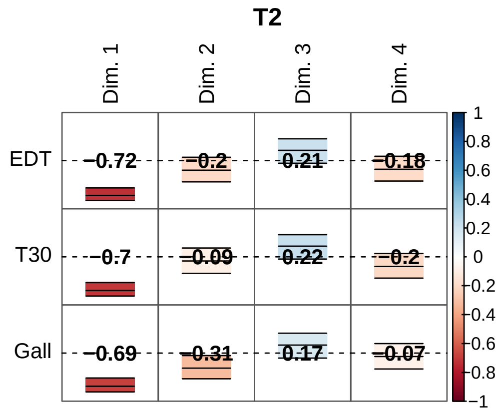
Player T2

The overall performance level (Dim. 1) of player T2 presents a strong significant negative correlation (approx. -0.7) with all the evaluated room acoustic parameters (see Fig. 6.8). The player mentioned during the interview that the performance tends to be louder in very dry rooms (see Tab. 6.7), which confirms the trends extracted from the signal analysis. In addition, a moderate negative correlation (-0.31) is found between the dynamic variations of the performance (Dim. 2) and G_{all}. The player seems to not be particularly influenced by acoustics regarding temporal aspects of the performance.

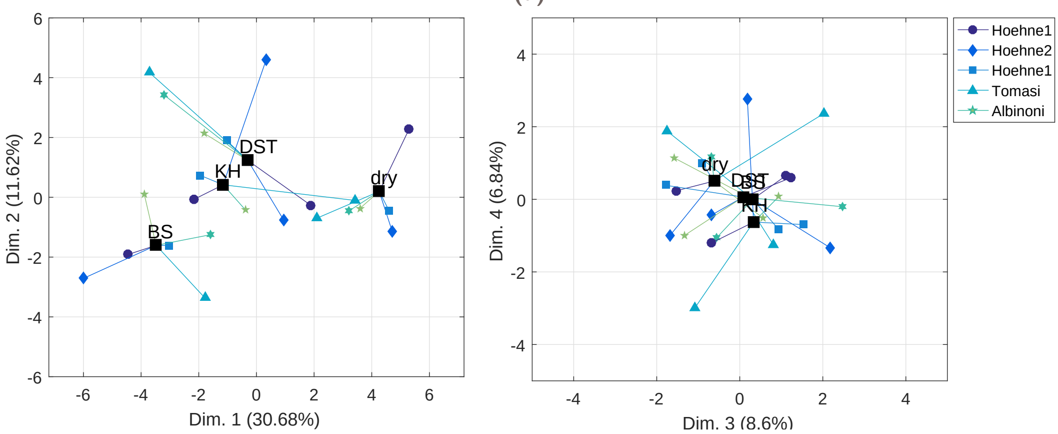
Analyzing the partial clouds (see Fig. 6.8), it is clear that the performances in *dry* present higher values in Dim. 1, with respect to the other rooms. Levels of dim. 1 and 2 of performances in rooms *DST* and *KH* are fairly close, suggesting that the performance in those rooms are similar in terms of musical dynamics. Performances recorded in room *BS* show lower levels of both dim. 1 and 2. The partial clouds referring to dim. 3 and 4 show that temporal aspects of the performances in different rooms are highly overlapped, confirming the absence of influence of room acoustics on temporal aspects of the performance.

Performance aspect	Adjustments
Tempo	More reverberation induces a "lighter" performance.
Dynamics	Higher reverberation leads to a softer performance. The performance is louder in a very dry room. Playing too loud in a room with many "bounces" (much reverberation) blurries the performance.
Articulation	The articulation is more staccato in rooms with longer reverberation, and notes are longer in dry rooms.
Expressivity	Performance in a dry room needs to be less expressive. The opposite for longer reverberation.
Other	The model of trompet is important. One must control the volume when playing piccolo trumpet. Used to play american trumpet and recently switched to german.

Tab. 6.7.: Interview responses of player T2



(a)



(b)

Fig. 6.8.: Results of player T2

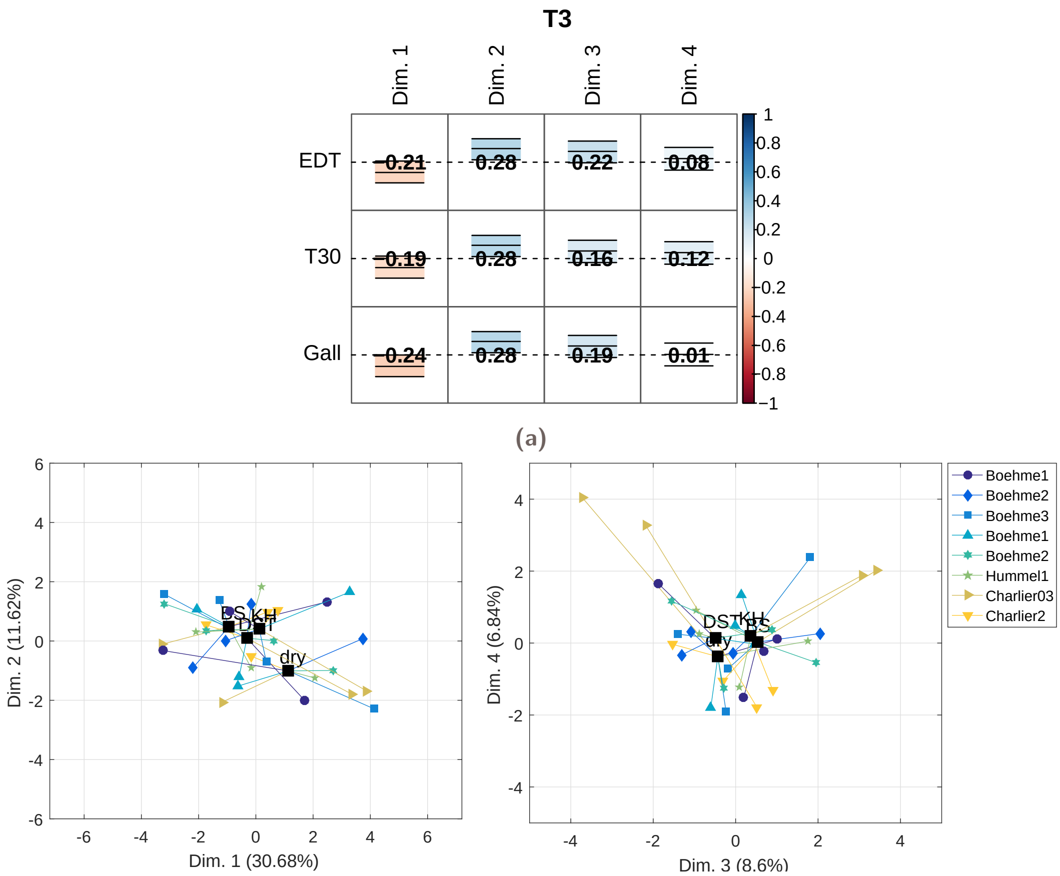
Player T3

The performances of player T3 (see Fig. 6.9) present weak correlations between musical dynamics (overall level and dynamic variations) and room parameters. However, only those corresponding to dynamic variations (dim. 2) are statistically significant, confirming the personal feedback provided by the player during the interview, stating that with more hall the dynamic differences are bigger. In addition, the player mentioned as well that with more hall the tempo tends to be reduced (see Tab. 6.8). However the correlations between Dim. 3 (*tempo*) and the room parameters are rather weak (close to 0.2). Finally, no significant correlations are found between Dim. 4 and any of the room parameters.

The same can be extracted from the partial clouds. However, it appears that only those performances recorded in room *dry* can be easily distinguished from the rest of room. In addition, there is a high variance among different pieces recorded in the same room e.g. see values of pieces Boehme 3 and Boehme 1 mapped on Dim. 1. This could either suggest that this player could adapt to room acoustics differently depending on the musical character of the piece, or the consistency in the playing style between consecutive recordings is low, presenting noticeable differences in every recording, regardless of the present acoustics.

Performance aspect	Adjustments
Tempo	"The tempo is not the same. With more hall I play slower."
Dynamics	"With more hall I make more dynamic differences." Playing piano passages in a big room one can let the air flow and feel more relaxed than in dry rooms.
Articulation	"With less hall I pay more attention to articulation". The reverb helps with legatto articulations. If the room is dry one needs to play longer notes.
Expressivity	"With more hall it is more fun to achieve more expressivity, but I always try to meet the expression".

Tab. 6.8.: Interview responses of player T3



(b)

Fig. 6.9.: Results of player T3

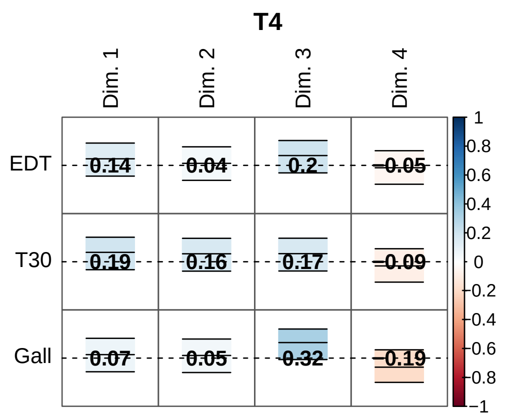
Player T4

The performances of player T4 do not present significant correlations between performance aspects and room acoustics (see Fig. 6.10). This suggest that this player does not follow a systematic approach regarding performance adaption under different room acoustics.

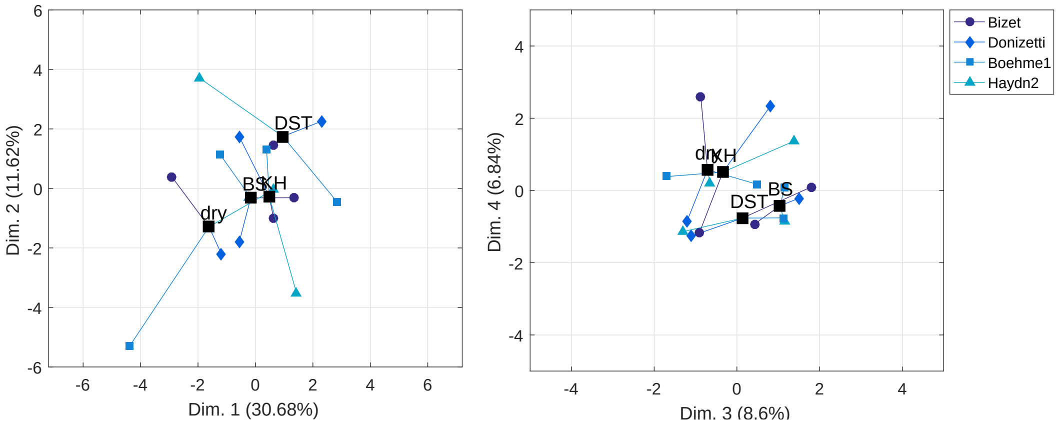
In addition, the performances recorded in room *dry* present lower values of Dim. 1 (performance level) and 2 (dynamic variations), which somehow contradicts a statement given during the interview, suggesting that if one does not hear oneself properly (e.g. in dry environments) it is necessary to play louder. The values mapped on Dim. 3 and 4 suggest that the effects on temporal aspects of the performance are rather small, confirming the player's statement that affirms that tempo is not particularly influenced by acoustics (see Tab. 6.9).

Performance aspect	Adjustments
Tempo	The tempo is not particularly influenced by acoustics
Dynamics	If one does not hear themself properly, it is necessary to play louder. Dynamics are easier to control with some reverberation. Dry acoustics create a "tinned" sound. The public in a room makes a difference. When a room is full, one must play a bit louder.
Articulation	The articulation is faster and more clear in a more reverberant room.
Expressivity	Playing with acoustics help to communicate with the public.

Tab. 6.9.: Interview responses of player T4



(a)

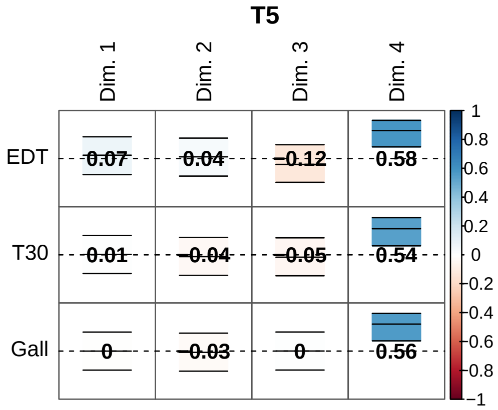


(b)

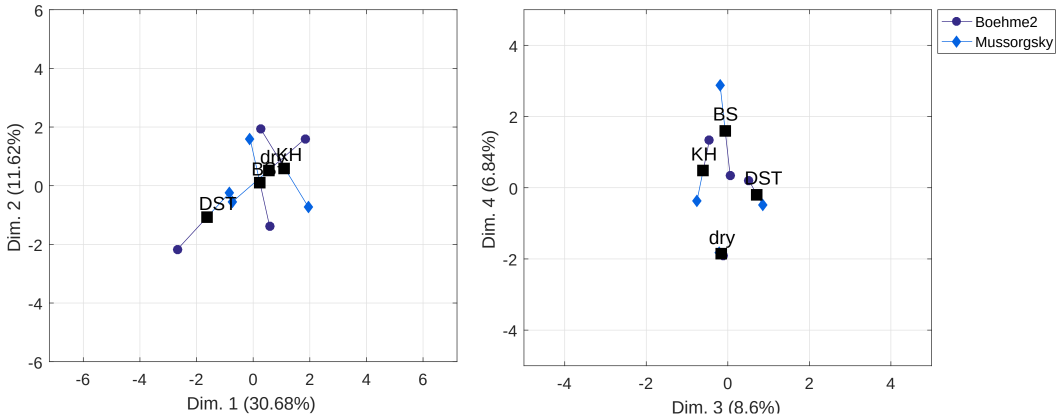
Fig. 6.10.: Results of player T4

Player T5

Player T5 (see Fig. 6.11) appears to adjust only performance aspects related to temporal variations. Hence, a moderate positive significant correlation is found between all room parameters and Dim. 4, suggesting that more reverberant environments lead to an increase of *tempo* variations. Increasing *tempo* variations is a musical resource often used to emphasize the phrasing and expressivity of a performance. The signal analysis results partially confirm the subjective response of the player, stating that the freedom and safety of good acoustics are beneficial for musical phrasing. Contrarily, the musician stated that performances in drier rooms result in an increase of overall level, which is however not confirmed by the signal analysis results. Finally, the lack of adjustment of the overall *tempo* partially agrees with the response of the musician, who says that it could be unconsciously affected, but there are no mentions of intentional or conscious variations (see Tab. 6.10).



(a)



(b)

Fig. 6.11.: Results of player T5

Observing the partial clouds confirms that the only identifiable effects of room acoustics on performance adjustments are present in Dim. 4 (*tempo* variations). The values corresponding to room *dry* are the lowest, as opposed to *BS*, while values of recordings in *KH* and *DST* are similar. This is particularly visible for the piece *Mussorgsky*, and again confirms the player statement regarding freedom and safety of acoustics supporting the musical phrasing.

Performance aspect	Adjustments
Tempo	The tempo could be unconsciously affected. If one does not feel comfortable there might be a feeling of rush to finish, increasing the tempo.
Dynamics	Playing in drier rooms results in louder performance because one can not hear the performance properly. The phrasing is more difficult in dry rooms because of focusing on technical aspects.
Articulation	Never considered if it changes, but it could be due to being more comfortable in some rooms.
Expressivity	Feeling freedom and safety of good acoustics help one with the phrasing and not being focused on technical aspects.
Other	Acoustics are particularly important for wind players because it is important to play relaxed, specially regarding the lips. Feeling relaxed is the most important aspect regarding to how acoustics influence the performance.

Tab. 6.10.: Interview responses of player T5

Player T6

Although a number of weak correlations are found between room acoustics and performance dimensions, only those regarding Dim. 2 and G_{all} are significant (see Fig. 6.12). This suggests that player T6 tends to increase the amount of dynamic variations in more environments with more overall acoustic energy. The subjective responses of the player are in partial agreement with those results (see Tab. 6.11), since *tempo* does not seem to be affected, and trends regarding the adjustment of dynamic variations are observed.

The partial clouds show a great variance of values in Dim. 1 among different pieces performed in the same room. However, it can be seen that performances in *BS* present a higher value in Dim. 2 than those recorded in the rest of the rooms. It is not straightforward to extract conclusions from the mappings of Dim. 3 and 4, which present great variability among all recordings and do not show clear trends.

Performance aspect	Adjustments
Tempo	The tempo is not much affected
Dynamics	"I try to adapt the performance e.g. not too loud in a large hall"
Articulation	"In more reverberant rooms the sound must be faded away faster, in dry rooms the notes should be longer"
Expressivity	"I think the expressivity should not depend on the acoustics"
Other	"I feel better in a good sounding room"

Tab. 6.11.: Interview responses of player T6

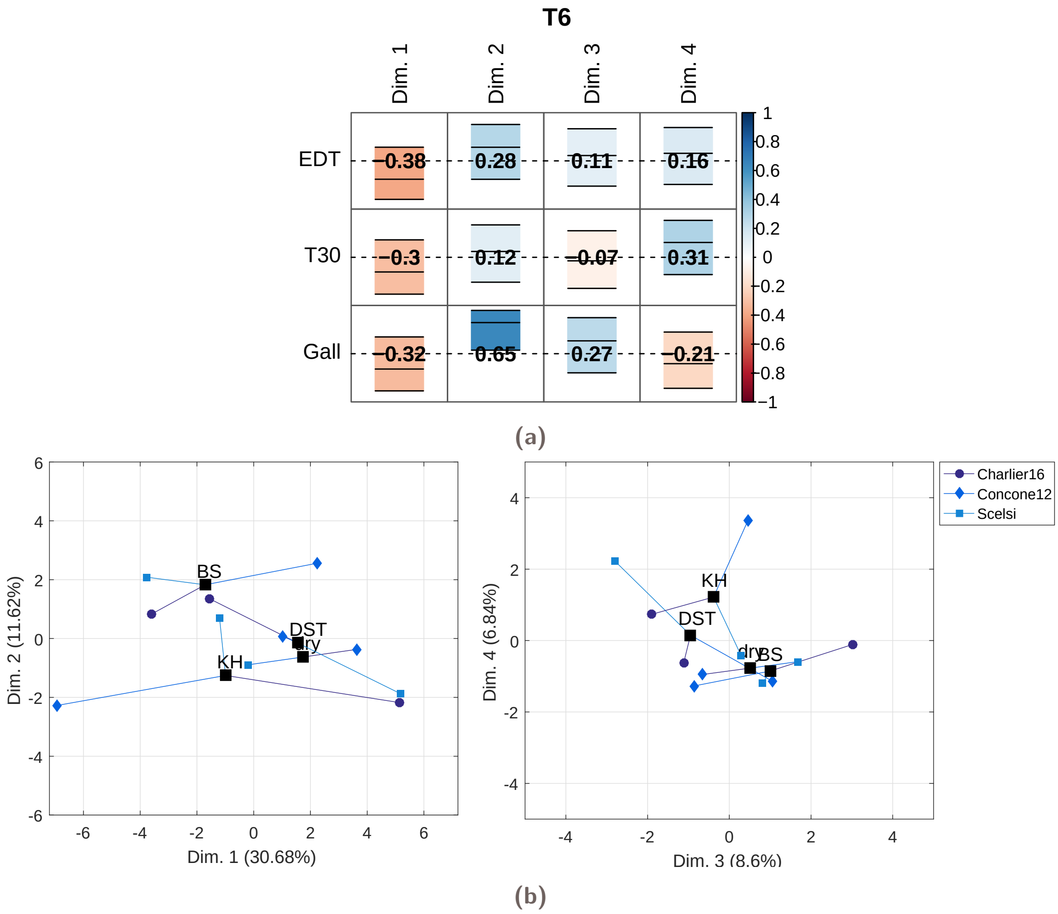


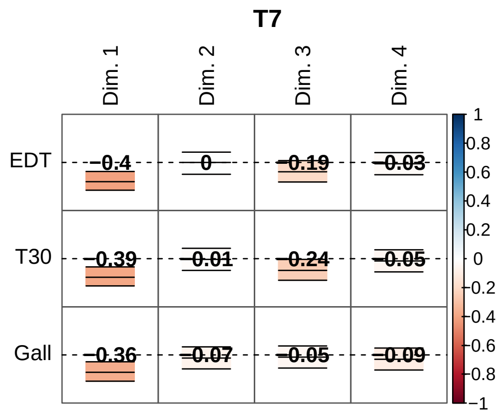
Fig. 6.12.: Results of player T6

Player T7

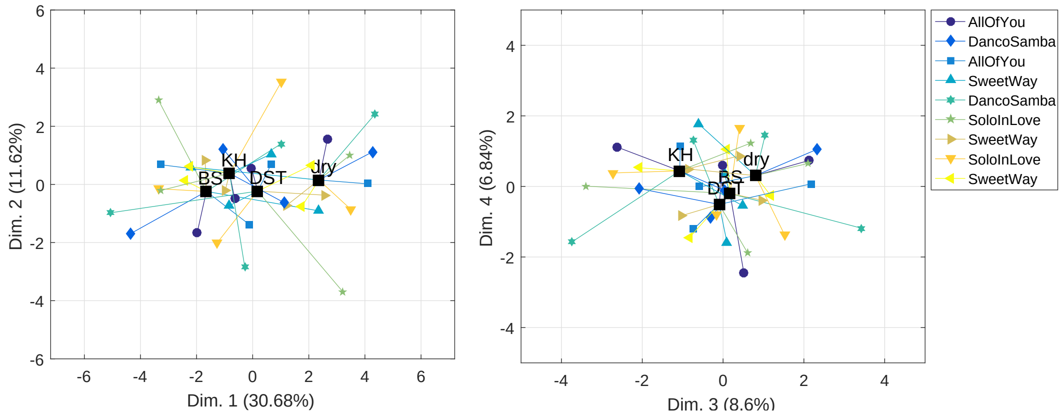
The main effects of room acoustics on the performance of player T7 (see Fig. 6.13) are related to overall performance level (Dim. 1) and overall *tempo* (Dim. 3). A statistically significant moderate negative correlation is found between all room parameters and Dim. 1, suggesting that the overall level of the performance decreases according to the increase in reverberation and room level, confirming the adjustment description given by the player regarding level and timbre adjustment. In addition, a weak negative correlation between Dim. 3 and reverberation parameters (EDT and T_{30}) suggest a tendency to slightly increase the overall *tempo* of the performance. However, the player stated that the main idea consisted on keeping the same *tempo* regardless of the acoustic conditions, or it could even be reduced in some cases (see Tab. 6.12). While the generalized impresion regarding *tempo* adjustment is that a longer reverberation time leads to a reduction of the tempo, it is important to note that player T7 is a jazz musician, and thus the performed repertoire had a different musical character, when compared to other players.

Performance aspect	Adjustments
Tempo	The player tries to keep the tempo, but it is possible that more reverberant rooms lead to a decrease. If the room is too big the tempo needs to be modified.
Dynamics	The performance is louder in a dry room. The musician does not feel it during when playing, it is somehow instinctive. The performance is perceived quiter in a dry room due to not hearing the "wind" (high frequencies).
Articulation	Information is lost in fast passages when playing in a big room, thus more difficult to control.
Expressivity	The room has an effect, a dry room is more "intimate". The sound being different in each room leads to different performances. The phrasing is better in a dry room, because reverberation hides the mistakes.
Other	"I visualize first the music in my mind and then translate it into the room" The timbre can change due to different mouth shapes changed unconsciously depending on the acoustics. The player always uses american trumpet.

Tab. 6.12.: Interview responses of player T7



(a)



(b)

Fig. 6.13.: Results of player T7

It can be seen in the partial clouds that the centroid of the room *dry* (overall level) presents the higher value mapped on Dim. 1, while all the rooms present a very similar values of

Dim. 2, suggesting that the adaption of level dynamics are not systematically linked to the acoustic conditions. In regard to Dim. 3 and 4, *KH*, the room with longest reverberation time, presents the lowest value of Dim. 3 (fastest *tempo*), as opposed to *dry*, which presents the slowest *tempo*. Trends are not easily identifiable in Dim. 4, since there is a great overlap among different pieces recorded in each room.

Player T8

The performances of player T8 present clear trends regarding dynamic variations (Dim. 2) and overall *tempo* (see Fig. 6.14). A strong negative correlation is found between Dim. 2 and all room parameters, specially with T_{30}, suggesting that an increase of reverberation leads to a reduction of dynamic variations i.e. a more constant sound level. In addition, longer reverberation times also lead to a decrease of overall *tempo*.

Although the correlations between Dim. 1 (overall level) and the room parameters are not significant, when observing the partial clouds it becomes evident that the player modifies the level of the performance systematically depending on the acoustics of the room, given that the centroids and partial results present fairly different values. In this sense, the overall level of *BS* and *DST* present the greatest difference, with *dry* and *KH* presenting similar values. In addition, the performances recorded in room *dry* present a significant higher value than those recorded in any other room, which are similar. The correlation analysis referring to Dim. 3 can be further analyzed by observing the centroids and individual recordings mapped on the partial clouds. It is clear that the performances in *dry* have a smaller value than those of the other rooms, which increase gradually according to the reverberation of the room. Finally, no clear trends are visible regarding *tempo* variations (Dim. 4).

Some of the information provided by the musician during the interviews is confirmed (see Tab. 6.13). For instance, the musician stated that the performance tends to be quieter in more reverberant rooms. Although this is not systematically observed (*KH* and *dry* present similar overall level values), it is clear that the level is systematically adjusted. In addition, the player also confirmed that that the tempo is increased in drier environments. Contrarily to that, the musician stated that more reverberation leads to a more expressive performance. Considering that expressivity is a mixture of musical variations, deviations from the exact [Sea37], one would expect an increase of dynamic variations. However, a clear trend is observed in the opposite direction, suggesting that the dynamic variations are reduced in more reverberant environments.

Performance aspect	Adjustments
Tempo	The tempo changes depending on the room. It is easy to get faster in very dry rooms.
Dynamics	The performance is quieter in more reverberant rooms.
Articulation	The articulation is more staccato with more reverberation.
Expressivity	More reverberation leads to a more expressive performance.
Other	The musician tries to adapt the performance according to the acoustic feedback of the room.

Tab. 6.13.: Interview responses of player T8

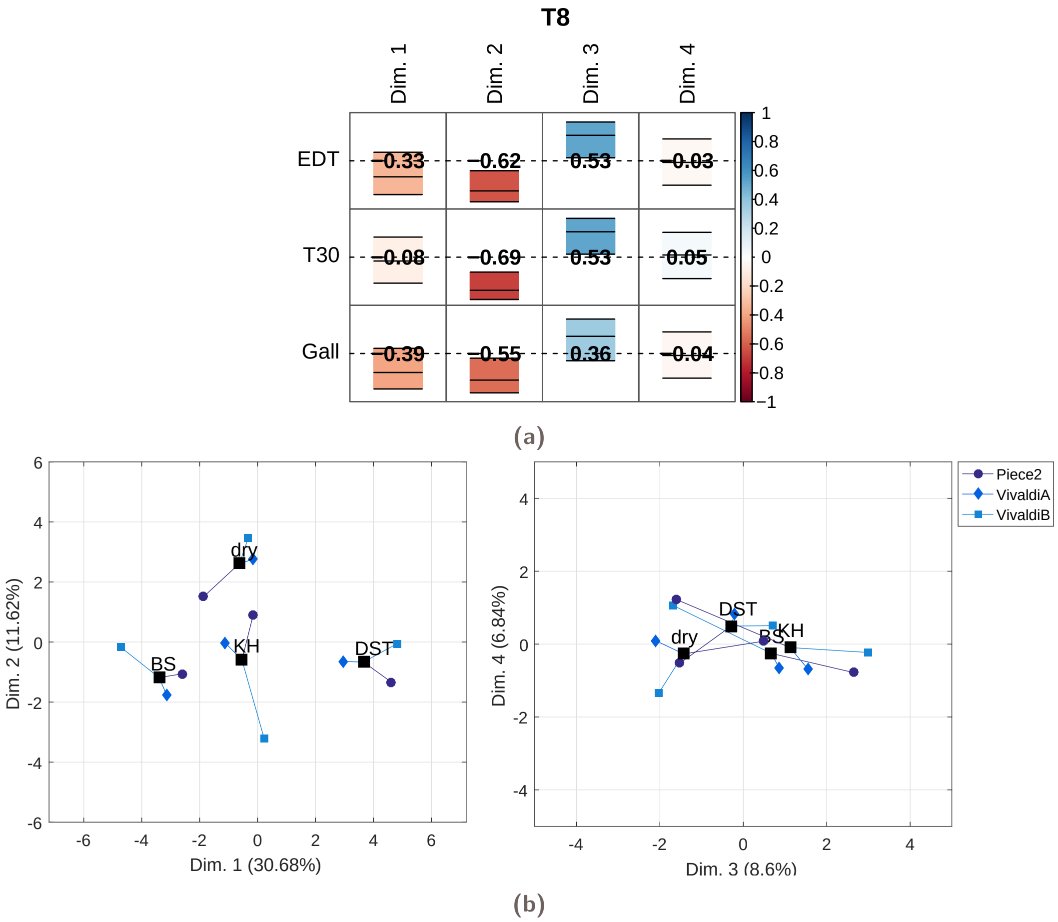


Fig. 6.14.: Results of player T8

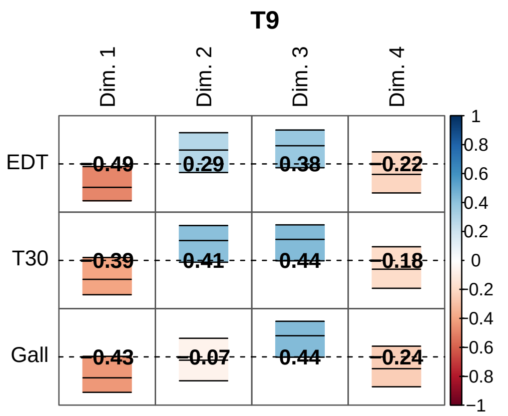
Player T9

Player T9 presents a moderate significant negative correlation between overall performance level (Dim. 1) and all the room acoustic parameters (see Fig. 6.15). This confirms that although the musician does not explicitly describe the implemented dynamic adjustments, during the interview it is mentioned that the performance is adjusted to compensate for the feedback received. In addition, Dim. 3 (overall *tempo*) is positively correlated with the parameters, meaning that more reverberant environments lead to a generalized decrease of level and *tempo*. In this case, the result contradicts the feedback provided by the musician, suggesting that the *tempo* changes could indeed be implemented unconsciously (see Tab. 6.14).

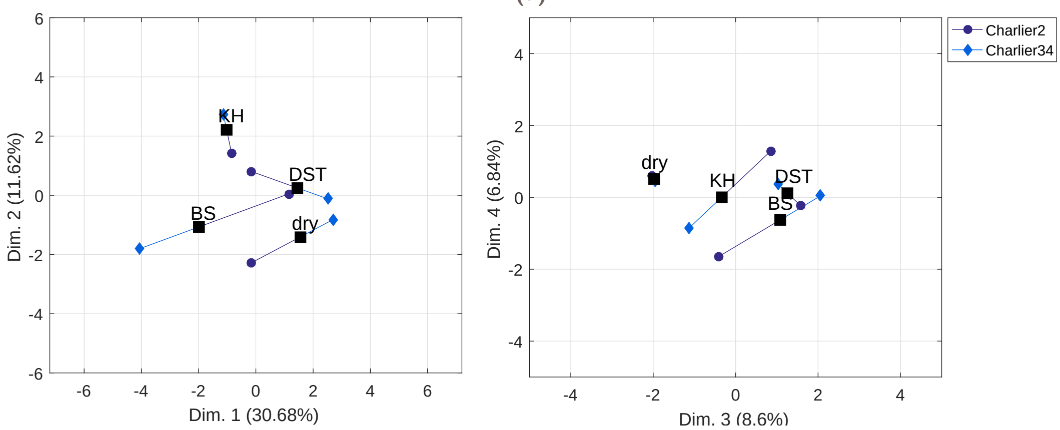
The partial clouds show that performances in *dry* or *DST* have a generally higher value in Dim. 1, while *KH* presents a higher value in Dim. 2. This suggests that performances recorded in *KH* present more dynamic variations. Regarding temporal aspects, those recordings in room *dry* are located in lower values of Dim. 3, as opposed to *DST* or *BS*. This means that performances in *dry* conditions are usually faster than those recorded in other rooms. There are no clear indicators of significant differences in *tempo* variations (Dim. 4).

Performance aspect	Adjustments
Tempo	The tempo is not changed.
Dynamics	Not sure about dynamic variations. A bigger room (in terms of dimensions) leads to a louder performance, to "fill up" the room. The musician would compensate for what the room gives.
Articulation	The articulation changes a bit.
Expressivity	Being more relaxed helps being more expressive.
Other	The most influenced aspect is the relaxation. The appropriate room provides a comfortable feeling. The perceived timbre is slightly darker in drier rooms, and slightly brighter when the feedback can be heard. The timbre is not consciously changed.

Tab. 6.14.: Interview responses of player T9



(a)



(b)

Fig. 6.15.: Results of player T9

Player T10

The most clear effect regarding performance adaption of player T10 are related to overall performance level (Dim. 1), which presents a significant negative correlation with all room parameters (see Fig. 6.16). The musician commented during the interview that the dynamics are indeed modified in order to achieve a certain sound, but there are no explicit indications regarding the adjustment of overall level (see Tab. 6.15). The overall *tempo* or *tempo* variations do not show significant changes, and they are in fact not mentioned in the interview.

Performance aspect	Adjustments
Tempo	There are not direct mentions to tempo changes.
Dynamics	The dynamics of the piece Cavadini1 were not much affected. In Cavadini2 the dynamic range depends on the feedback of the room, which is needed to control the dynamics. The dynamics are modified trying to achieve a certain sound. One must play louder in bigger rooms (bigger not necessarily related to acoustics).
Expressivity	While playing a slow piece the room determined how much the player felt could "open" the performance.
Other	Some feedback is needed to play, but not too long.

Tab. 6.15.: Interview responses of player T10

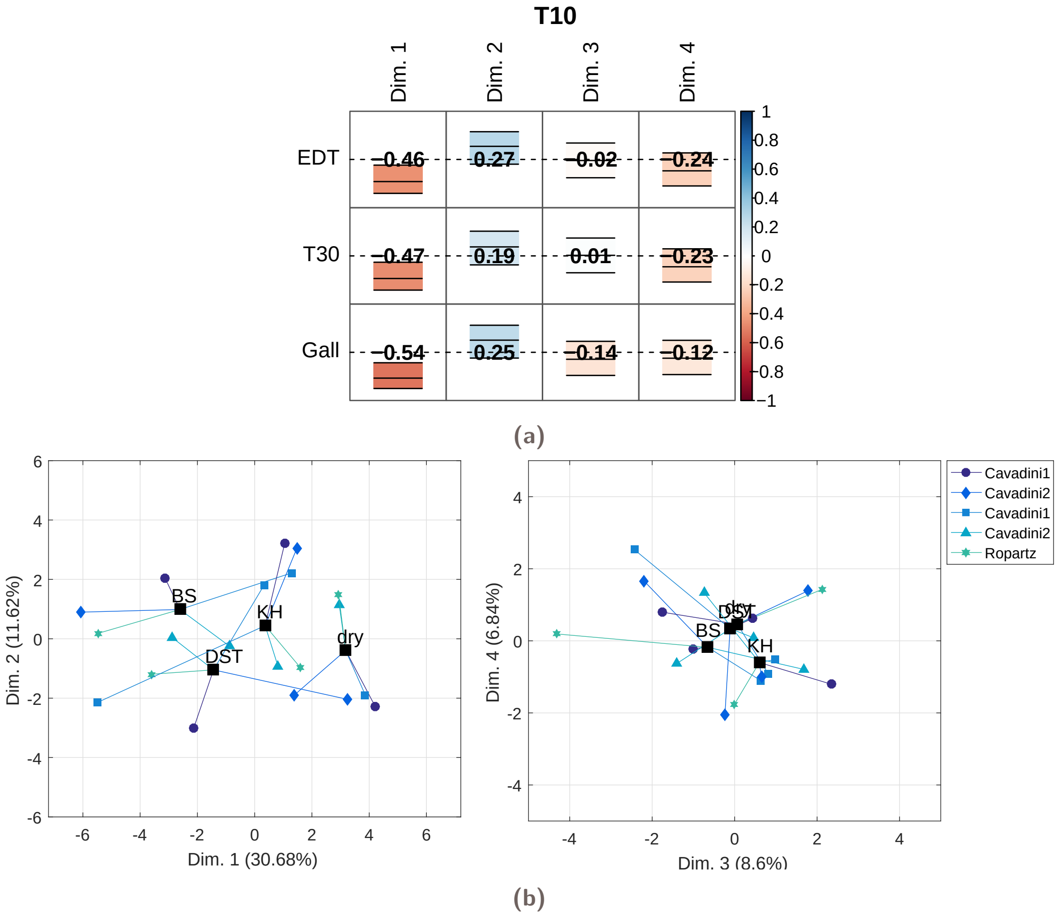


Fig. 6.16.: Results of player T10

The partial clouds show that the values of Dim. 1 corresponding to performances recorded in the room *dry* are generally higher than those of the other rooms. The rest of the dimensions show highly overlapped values, suggesting that there are not clear modifications regarding those performance aspects when performing in different acoustic conditions. The player explicitly mentioned that the dynamics of the piece *Cavadini1* are not particularly affected, while dynamic range is actively adjusted in *Cavadini2*, depending on the room feedback. Nevertheless, this is not clearly observed in the partial clouds.

Player T11

The only significant performance adjustments implemented by player T11 are a decrease of overall level performance (Dim. 1) when performing in environments with longer reverberation (see Fig. 6.17). Moderate significant correlations are found between Dim. 1 and reverberation parameters (EDT and T_{30}). The rest of the dimensions do not seem to be particularly affected, although a moderate non significant correlation is found between Dim. 3 (overall *tempo*) and total energy of the room (G_{all}). The feedback provided by the musician in the interview does not present clear relationships with the performance analysis, since the mentioned adjusted aspects (dynamic variations and *tempo* variations) do not present significant correlations with the room parameters. However, the lack of overall *tempo* adjustment appears to confirm the subjective comments, which state that overall *tempo* is intended to be kept constant regardless of the acoustic conditions (see Fig. 6.16).

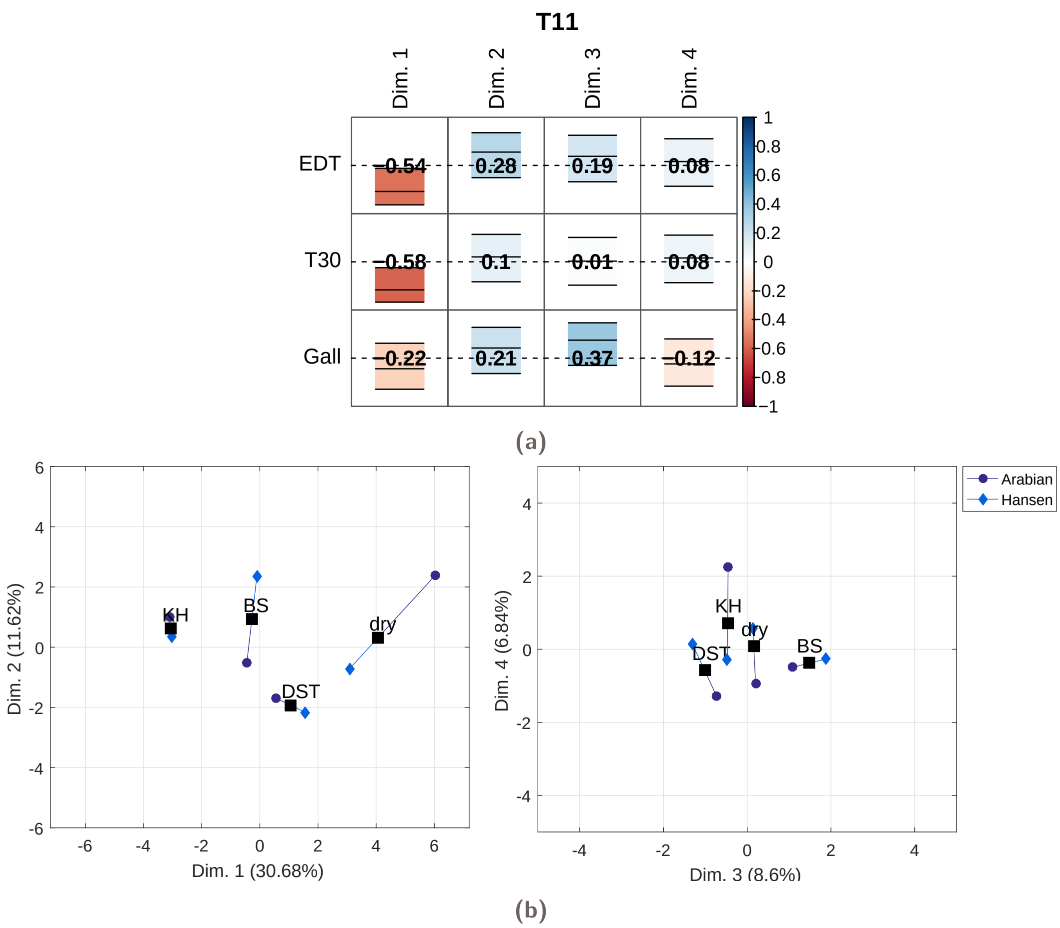


Fig. 6.17.: Results of player T11

The partial clouds show that performances recorded in each room present similar values of Dim. 1, which decrease progressively with higher reverberation times (*dry, DST, BS, KH*). Performances in room *DST* seem to present smaller dynamic variations than the rest of the rooms. Finally, the overall *tempo* (Dim. 3) seems to decrease in room *BS*, and the *tempo* variations (Dim. 4) seem to be similar in all the rooms.

Performance aspect	Adjustments
Tempo	More tempo changes in more reverberant environments The overall tempo is intended to be the same regardless of the room, but it could be faster in more reverberant rooms.
Dynamics	More dynamic variations in more reverberant environments
Expressivity	Better acoustics help to relax, and overall to a more expressive (more dynamics, tempo variations) performance.
Other	The timbre is more brilliant in a hall.

Tab. 6.16.: Interview responses of player T11

6.1.5 Summary of results

A graph including the correlation values of performance dimensions 1 to 4 and the room acoustic parameters corresponding to the rooms included in the experiments is depicted in Fig. 6.18. This graph summarizes the information presented in section 6.1.4.

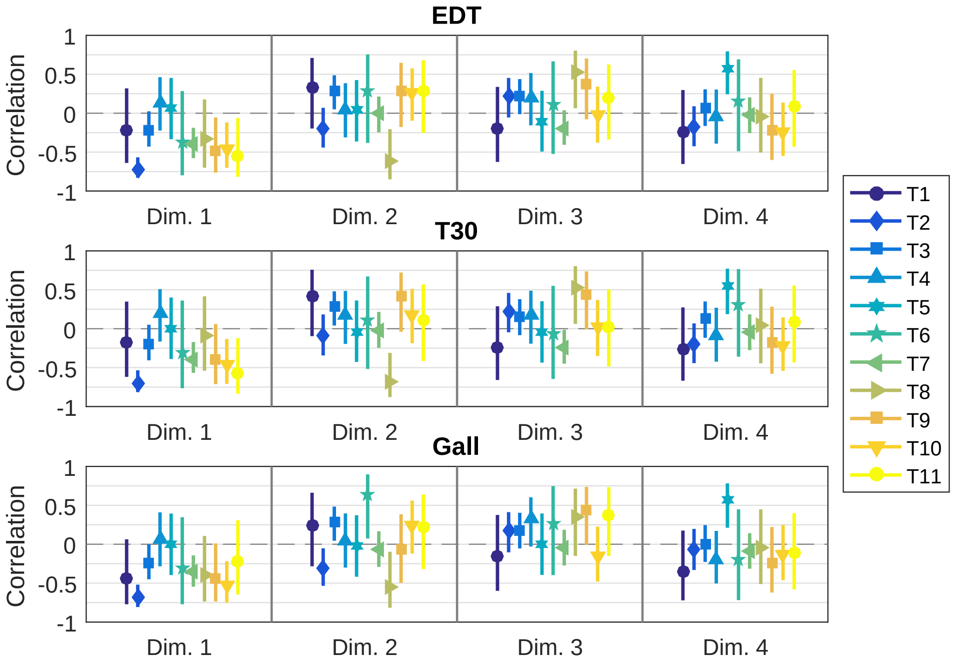


Fig. 6.18.: Correlation between performance dimensions and room acoustic parameters. Vertical bars indicate the 95% confidence intervals of the correlation values.

The following conclusions can be extracted from the summarized results:

- There is a generalized tendency to decrease the performance level in more energetic and reverberant rooms, leading as well to a darker timbre. None of the players significantly increases the level of the performance.

- The adjustment of dynamic variations in different acoustic conditions seem to be largely individual. While most of the players do not present statistically significant adjustments, players T2, and T8 tend to reduce the dynamic variations in more reverberant or energetic rooms, and players T3 and T6 increase the dynamic variations in more energetic environments.

- Although a generalized trend towards a reduction of the overall *tempo* in more reverberant or energetic environments is observed, only players T8 and T9 present statistically significant adjustments. Contrarily, player T7 implements a slight overall *tempo* increase in rooms with longer reverberation time.

- No significant trends are observed for the majority of the players regarding the adjustment of *tempo* variations. Only player T5 presents an increase in *tempo* variations when performing in more reverberant or energetic environments.

The computed correlation values between room acoustic parameters and performance dimensions can be used to generate clusters of players that share similar adaptation strategies. The Euclidean distance between the correlation values presented in Fig. 6.18 is used to obtain a relative distance between all the players. Then, a tree of hierarchical clusters is generated. This process is repeated for every dimension, obtaining a partial clustering for each performance aspect. The resulting clusters are presented in Fig. 6.19.

The players' behavior regarding Dim. 1 (overall level) can be classified into two main clusters: those who moderately or strongly adjust their performance (most of the players), and those who do not present significant differences (players T4 and T5). Among those players who implement performance changes, player T2 presents a much higher correlation than the others, constituting a separated (sub)cluster.

The grouping seems less clear regarding Dim. 2 (dynamic variations). In this case, most of the players do not present statistically significant correlations, thus forming a main cluster. Players that systematically adjust the dynamic variations present opposed behaviors, thus forming different clusters.

Three main clusters are observed regarding the overall *tempo* adjustment: those players who significantly decrease it in more reverberant or energetic environments (players T8 and T9), those who tend to decrease it but the result is not statistically significant (T2, T3, T4, T6, T11), and those who present slight non significant tempo increase (T1, T5, T7, T10).

Only one of the players (T5) systematically adjusts the *tempo* variations in the performance, representing a cluster. The rest of the players constitute another big cluster.

In order to obtain a unique clustering accounting for all the performance adjustments, the hierarchical clustering process is applied to the data containing the correlation of all

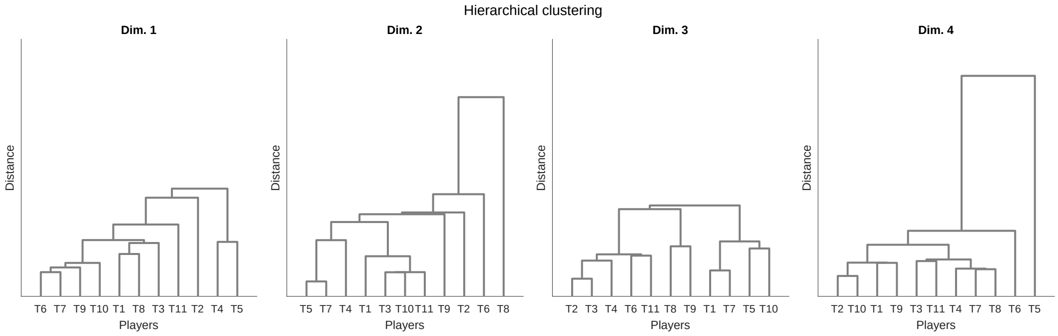


Fig. 6.19.: Hierarchical clustering of player behavior in each dimension.

dimensions. In this case, the relative distance ratings are weighted using the explained variance of the first four performance dimensions (0.31, 0.12, 0.9, and 0.7, respectively).

The resulting clusters and the relative distances between all players are presented in Fig. 6.20. The hierarchical clusters show a big cluster composed of players most of the players (T1, T3, T4, T5, T6, T7, T10, T11), another cluster of two players (T2, T9) and one player that does not correspond to any of them (T8). By comparing the clustering result with the correlation data from Fig. 6.18, it is possible that player T8 alone represents a cluster due to the fact that their behavior regarding Dim. 2 is opposed to the rest of the players, thus being a differentiating characteristic. However, given the multidimensionality of the data, it is not straightforward to interpret the relationships between different clusters and its members.

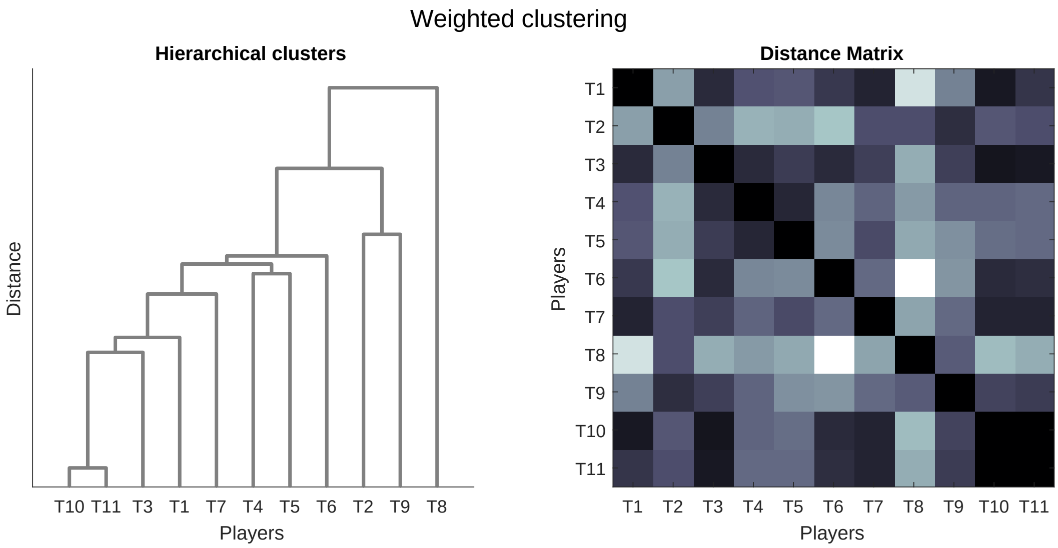


Fig. 6.20.: Hierarchical clustering and relative distance matrix of player behavior weighted by the variance of the four first performance dimension. Brighter cells in the distance matrix denote higher distance between players.

6.1.6 Discussion

Analyzing the feedback provided by the musicians during the interviews, one can easily conclude that musicians consciously adjust their musical performance according to the room acoustic conditions. Although not all the players present statistically significant correlations between the studied performance aspects and the selected room acoustic parameters, all the performers present systematic differences among performances recorded in different rooms, in at least one of the performance aspects. This would suggest that in some cases the relationships between the adjustments and the acoustics do not follow a linear trend, but are still present. However, given that the number of studied rooms in this investigation is rather limited, in order to unravel more complex relationships it is necessary to extend the set of rooms.

The main adjustment, shared by most of the players refers to a decrease of the overall performance level and timbre brightness when performing in rooms with longer reverberation times and more overall acoustic energy. The adjustments regarding other performance dimensions seem to be more individual, and possibly depending on the musical nature of the recorded pieces.

Although all musicians seem to be aware of the necessity of adjusting their performance to accommodate room acoustics, some of them explicitly mentioned that the changes could be in some cases unconscious. Thus, rather than objectively analyzing the separated aspects that constitute a musical performance and adapting them accordingly, the performance adaptation would be a result of an intuitive process, governed by the performer's subjective musical concept expression and the performance as a whole. In some cases, musicians' subjective perception regarding the performance adjustments could differ from those actually implemented, leading to a discrepancy between the verbal feedback and the automatized performance analysis. This suggests that the effect of room acoustics on live performance could imply an associated degree of intentionality, a subjective impression of modifying the performance without actually implementing any change.

When describing room acoustics, musicians tend to use simplified vocabulary, referring only to words such as "dry room", "reverberation" or "hall". It is common as well to describe the acoustical characteristics by dimension terms, such as "big room" or "small room". Although these terms can be often used to actually describe the physical dimensions of the room, it is also common to associate big rooms to concert hall acoustics, as opposed to small rooms with rather dry acoustics. For this reason, when conducting research including the verbal feedback of musicians with regards to room acoustics, it is important to clarify which specific aspects of sound are involved in their descriptions, in order to properly identify the possible relationships between the described performance and musical aspects and specific characteristics of the acoustics of the room.

6.1.7 Further work

The implementation of a MFA allowed the reduction of multidimensional data into a reduced set of 4 dimensions describing important aspects of the musical performance e.g. overall level, dynamic variations, overall *tempo*, and *tempo* variations. However, the perception of musical characteristics is a topic under research at the moment of writing this work and perception models of musical performance are not widely available for application. While the extracted MFA dimensions are strongly correlated to low and mid level features that describe perceptual and musical characteristics e.g. LUFS is indeed correlated to the perceived loudness of a signal, or BPM objectively describes the *tempo* of a performance - it is not clear to which extent they are relevant or appropriately describe the perception of subtle performance variations. To solve this issue it is necessary to conduct perceptual investigations using an extensive dataset of musical recordings with distinct performance characteristics. The dataset created during this project can serve as a starting point for the generation of these perceptual models, and preliminary work relating the MFA dimensions and the subjective musical perception is presented in Chapter 7.

The goal of the described experiments is to characterize and categorize the general behaviors of different players, while emphasizing the presence of individual performance strategies. To this end it was necessary to obtain a relatively large dataset, which resulted in approximately 400 recordings (364 after discarding recordings with perceivable errors and artifacts). Unfortunately, the size of the dataset does not allow for detailed analysis of every single recording, but a reduction of the data and dimensionality was necessary in order to implement a general and compact analysis. This means that aspects such as dynamic variations or *tempo* variations, which possess intrinsic temporal characteristics needed to be reduced to more simple descriptors, thus resulting in a possible loss of information. Given that it is probable that the musical character of the played pieces influences the nature of the performance adjustments, a detailed analysis of *tempo* curves and signal envelopes could provide useful information regarding specific musical characteristics that are systematically adjusted depending on the room acoustic conditions.

Finally, the information extracted from the results regarding performance adjustments, together with the implementation of perceptual models, can be used to generate performance synthesis models that account for the influence of room acoustics on live performance, and implement perceptually relevant modifications on the synthesized performances.

6.2 Organ performance in enhanced rooms

The history of organ music is tightly related to religious celebrations, and traditionally, the performances were held in religious spaces such as churches or temples. The acoustical properties of these spaces are characterized by very long reverberation times and low clarity. It is common for the reverberation time at mid frequencies to be greater than 5 s. However, during the last century it has become more common to install pipe organs in concert halls and performance rooms that present much lower reverberation times. In addition, acoustical conditions of small churches are often quite different from those of large cathedrals. Furthermore, given that the amount of permanent absorption in churches is normally rather small, the presence of audience impacts significantly on the acoustics. This means that organ musicians need to be able to accommodate a wide variety of acoustic spaces and perform pieces that were initially composed to be performed in a particular kind of acoustics.

In order to investigate the performance adaption strategies of semi-professional organ performers, a series of experiments were conducted using a room acoustic enhancement system (RAES), namely *Vivace*, in the Detmold Konzerthaus, allowing the modification of the acoustics of the hall in real-time. The work presented in this section is the result of a collaboration of the author with Dr.-Ing. Winfried Lachenmayr, doctoral candidate at the Detmold University of Music at the time of the investigations. The tests were conducted jointly, whereas the interviews were conducted by W. Lachenmayr and the implementation of the analysis algorithms, performance analysis, synthesis of results and interpretation was completed by the present author.

6.2.1 Technical set-up

A room acoustic enhancement system (RAES), namely *Vivace*, was temporary installed in the Detmold Konzerthaus for the conduction of organ performance experiments using variable acoustics. The operation of *Vivace* is based on the convolution of RIR with the sound captured in the hall by multiple microphones and playing back the convolved sound using a loudspeaker array. This results on an extended energy decay that blends the real acoustics of the hall with the electronically generated reverberation.

The system installed in the Detmold Konzerthaus was composed by two layers: a IOSONO Wave Field Synthesis (WFS) system, which is permanently installed, and the *Vivace* system. The WFS system controls a large array composed of 328 individual channels, and divided into a rectangular array of multi-actuator panels (MAP) surrounding the hall and a set of discrete loudspeakers placed on the ceiling of the hall. These arrays were used to generate a discrete array of 56 virtual sources that reproduced the reverberation electronically generated by *Vivace*. Both systems are commercial products and their internal algorithms are not publicly available, thus they can be regarded as black box systems.

The sound generated by the organ is captured by four directional condenser microphones. Two Neumann MK800 with super-cardioid directivity are placed at approximately 5 m height and 3.5 m distance from the organ, with the direction of maximum sensitivity pointed towards the organ. The other two microphones hang from the ceiling at approximately 6 m height and 5 m from the organ. An overview of the set-up and a close view of the organ console are displayed in Fig. 6.21.

Two different enhanced acoustics are designed, *soft increase* and *strong increase*, presenting an increase of reverberation time (RT_{30}) of 0.5 s, and 1.5 s, respectively, compared to the natural acoustics of the unoccupied hall, which is approximately 1.6 s (see Fig. 6.22). The reverberation time has been measured in third octave bands resolution by means of a hand held acoustic analyzer (NTi XL2) using the interruption method. The excitation of the room is done by pressing and releasing multiple keys of the organ simultaneously, ensuring that the measurement conditions are comparable to organ playing conditions. The response is measured at the player position and the measurement is averaged several times.

The organ performances are recorded using a MIDI port built into the manuals and pedal. Note on, note off, velocity and pitch messages are transmitted over the MIDI interface, and two separate channels are used to transmit the information of the manuals and the pedals. The MIDI data is recorded using a laptop with an audio interface. Simultaneously, binaural audio recordings are held using an artificial head placed 50 cm behind the head of the musician.

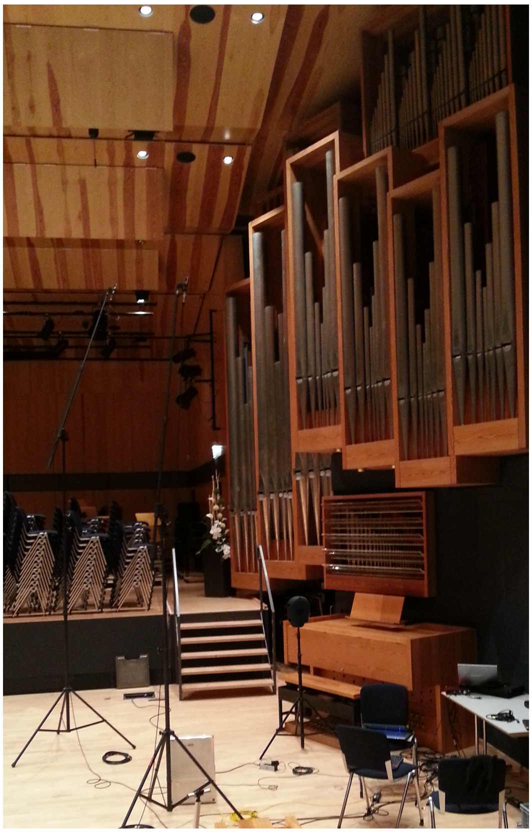


Fig. 6.21.: Overview of the organ experimental set-up.

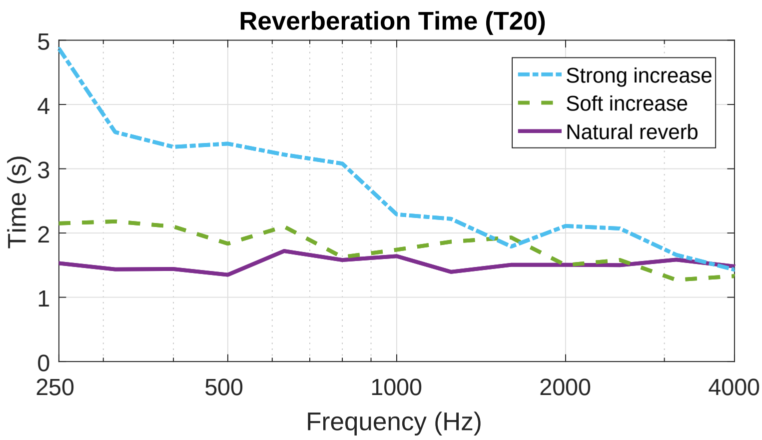


Fig. 6.22.: Reverberation time (RT_{30}) of the acoustic conditions tested in experiments.

Participant	Level
P1	Master (finished)
P2	Bachelor (5th semester)
P3	Bachelor (2nd semester)
P4	Bachelor (3rd semester)
P5	Bachelor (complementary instrument)

Tab. 6.17.: Participants data.

6.2.2 Experiment description

The goal of the experiment is to investigate the adaptation strategies of organ players during performance in different acoustic conditions. The main task is to perform these pieces and record them under different acoustic conditions. The MIDI recordings are then analyzed to extract relevant performance information.

The experiment is divided into two sessions which are completed within a time span of 7 months between each session, in order to compare the evolution of their performance strategies after a period of training. The sessions were completed during February 2015 and September 2015. In both sessions musicians are asked to prepare two or three short excerpts with different musical character. The participants are five bachelor and master students from the Detmold University of Music.

The first session consists of two parts, namely *blind* and *bon-blind*. During the *blind* test the musician is not provided with any previous knowledge about the acoustic conditions, which are changed randomly. In addition, they are not allowed to test the acoustics before the recording, thus they are required to adapt to the acoustic conditions during the actual recording. In the *non-blind* part, musicians are notified about a change in acoustics and they are allowed a short training period before recording, but they are not notified explicitly about the differences in the acoustic conditions.

The second session is likewise divided into two parts. In this case, there is a *non-blind* test and a *delay* test. During the *delay* test, the natural acoustic conditions of the hall are not modified, instead a delay between the instrument interaction and the generation of sound is included, imitating the effect of a large distance between the organ console and the pipes.

A summary of the data regarding the participants is included in Table 6.17. The details of the performed pieces is included in Tab. E.1 from Appendix E.

6.2.3 Results

The dynamics characteristics of a pipe organ remain constant during a performance, thus a player is not able to implement dynamic variations. For this reason, the performance aspects studied in these experiments are overall tempo and median note duration. The analysis of the median note duration is preferred of the average note duration due to the fact that in a typical performance the distribution of note durations will likely be skewed and not normal, and thus the median duration is more suitable when investigating changes in those

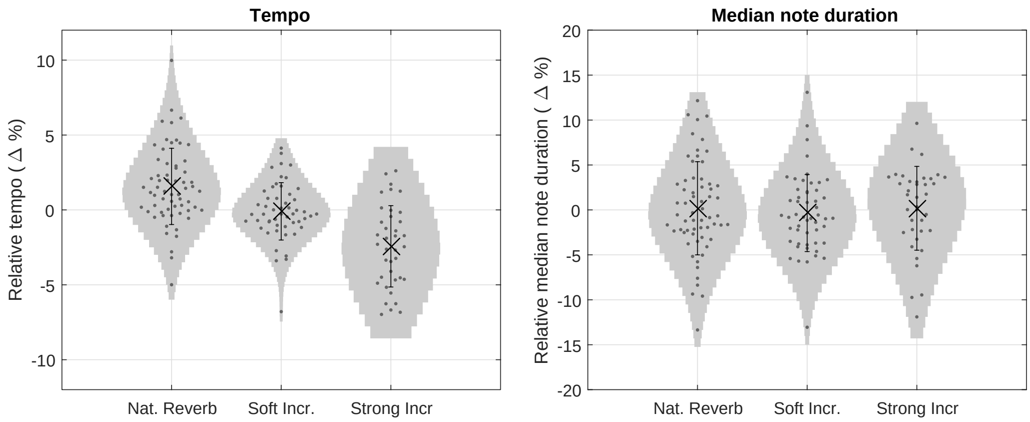


Fig. 6.23.: Normalized tempo and median note duration of all recorded performances. Grey areas represent a smoothened histogram using a normal kernel, point markers represent individual recordings, cross markers represent mean values and vertical bars represent standard deviation.

distributions. The duration of the notes is related to the playing articulation, as longer notes indicate a more *legato* articulation, whereas shorter notes suggest a more detached or *staccato* style.

An overview of the results including all the players, pieces and playing conditions is presented first, in order to identify general trends. However, given that it is expected that the performance changes are largely individual and depend on the musical piece and experimental conditions, a detailed analysis considering those effects is presented later.

General Trends

To investigate general trends, the extracted features from every recorded performance have been normalized with respect to their group. A group of performances is defined as those corresponding to the same player and piece.

$$\Delta Feat_{pl,pi,ac}(\%) = \left(\frac{Feat_{pl,pi,ac}}{\frac{1}{N}\sum_{n=1}^{N} Feat_{pl,pi}} - 1 \right) \cdot 100 \tag{6.2}$$

where $Feat$ corresponds to the studied performance feature, and pl, pi, and ac correspond to the player, piece and acoustic condition of the recording.

Once the normalized features have been calculated, they are organized in three groups, according to the acoustic conditions active during the recording, *Natural Reverb*, *Soft Increase*, and *Strong Increase*. A Lilliefors normality test performed on every group of normalized features confirms that all the groups present a normal distribution. Finally, a normal distribution estimate is fitted on the normalized data. The results of overall tempo and median note duration are presented in Fig. 6.23.

Some general trends can be extracted from the results:

- The average overall tempo is inversely correlated with the length of the reverberation. The average value of overall tempo is approximately 4% lower in *Strong Increase* conditions, compared to *Natural Reverb*. These changes are statistically significant with a *p-value* < 0.01.

- The average overall tempo values present a greater standard deviation in the condition *Strong Increase*, as compared to the other conditions.

- There are no statistically significant differences in average or standard deviation in median note duration.

In spite of the identification of general trends, it is worth noting that not all players contributed to the experiments with the same amount of recordings, thus it is expected that the impact of players with more recordings is greater in the general results than the players with fewer takes. In addition, taking into account previous investigations [SK15] it is expected that the performance changes depend on individual players and pieces, and valuable information can be extracted if an individualized analysis is applied on every group.

Overall Tempo

The overall tempo of the all recorded performances is shown in Fig. 6.24. The results are organized by musical piece, and the different players performing the same piece are shown together in the same graph. The resulting *p-value* from one way ANOVA tests performed on the results are detailed in Tab. 6.18.

As can be extracted from the results:

- Although the overall tempo of every player is usually significantly different from other players, the tempo adaption strategy seems to be partially common among all of them and different in every musical piece.

- All the players show a significant decrease of overall tempo performance for the piece *MendelssohnA* in at least one of the experimental conditions.

- Players P1 and P3 do not present statistically significant tempo variations when performing *MendelssohnA* under condition *blind*. However, then the players are explicitly notified about the acoustic conditions or allowed a time of training, they present a significant tempo reduction with increased reverberation.

- Although player P2 plays the piece at significantly different tempi in Exp1 and Exp2, their adaption strategy against increased reverberation time consists of a significant tempo reduction in all experimental conditions.

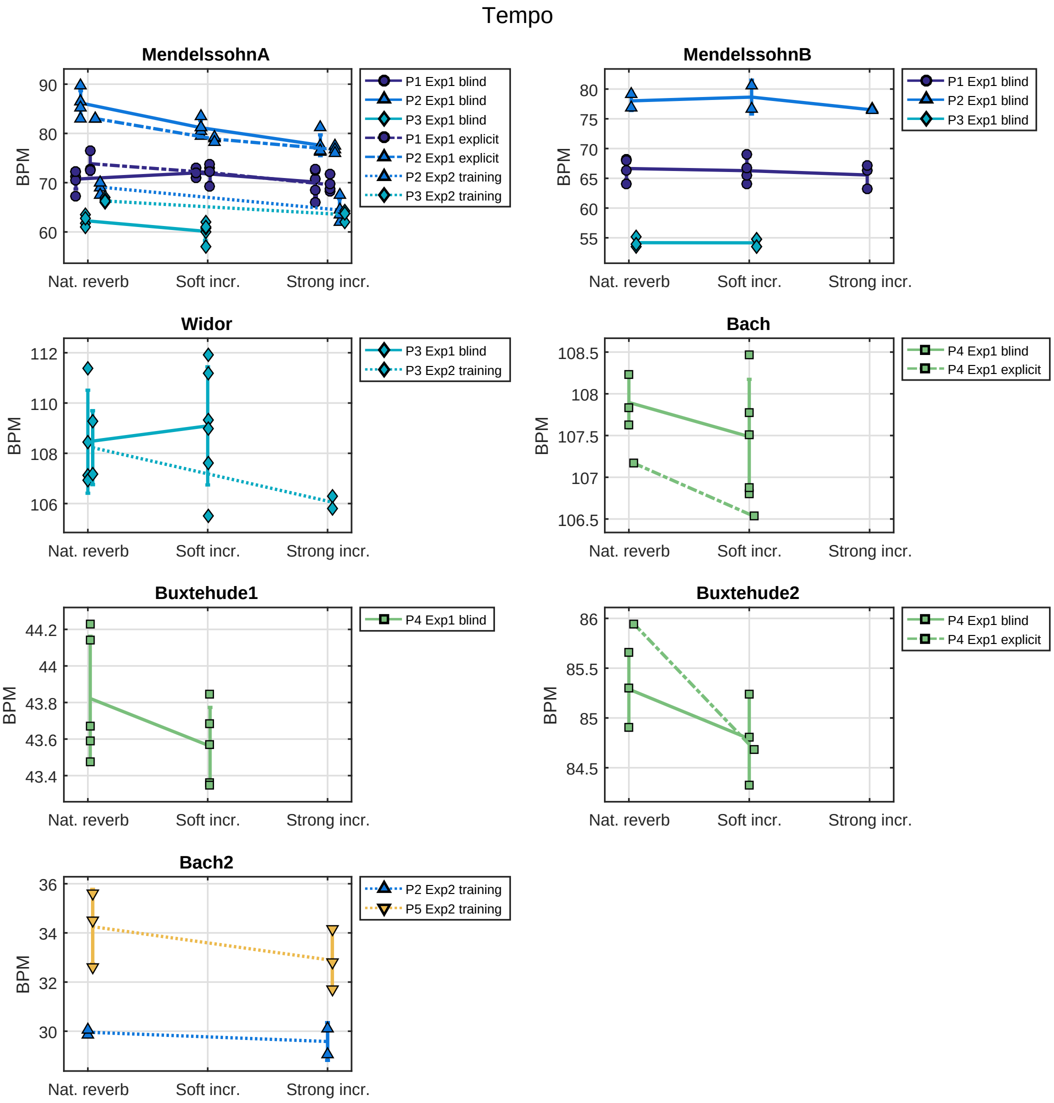


Fig. 6.24.: Overall tempo of the organ performances under different reverberation time.

- No significant tempo variations are present in the performances of any of the players of the piece *MendelssohnB*.

- The player P3 does not present statistically significant tempo changes in the piece *Widor* under *blind conditions*. However, when allowed a time of training before the recording the tempo reduction appears to be more consistent and closer to statistical significance.

- The player P4 does not present statistically significant tempo changes in any of the performances.

- In none of the performances statistically significant tempo increases are found with increased reverberation.

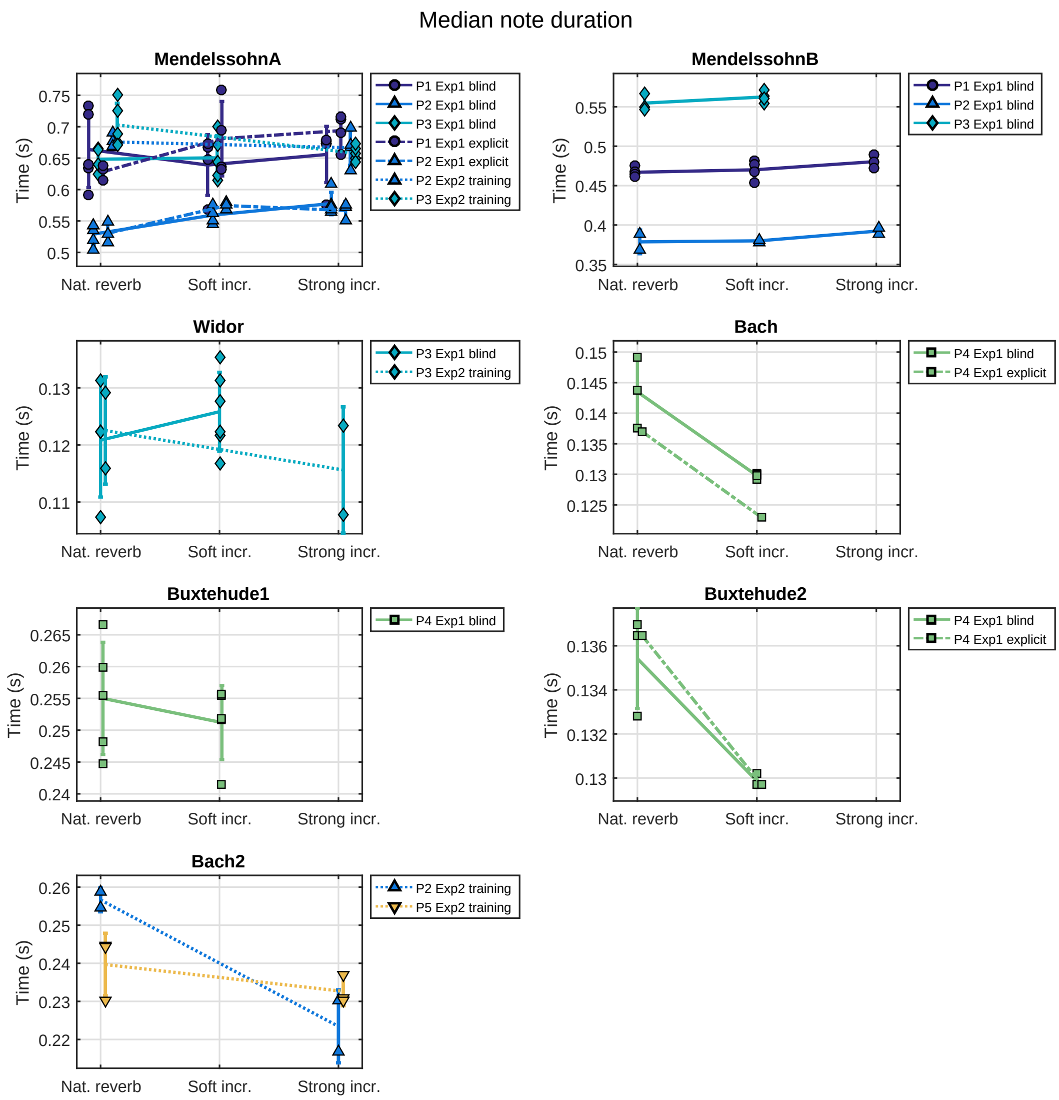


Fig. 6.25.: Median note duration of the organ performances under different reverberation time.

Note duration

The median note duration of all the recorded performances is presented in Fig. 6.25. As with the overall tempo, the results are separated by piece and all players are shown in the same graph. The *p-values* resulting from a one way ANOVA test are collected in Tab. 6.18.

The results suggest that:

- Players tend to exhibit different behaviors when performing *MendelssohnA*. Only P2 and P3 present statistically significant articulation changes. During the first session (Exp1), player P2 consistently increases the duration of notes (more *legato* articulation) when presented with longer reverberation times, and during the second session (Exp2) no significant changes are observed. Player P3 consistently reduces the duration of notes when allowed a period of training (Exp2), and no changes are observed in *blind* conditions.

Fig. 6.26.: First bars of the piece *MendelssohnA*.

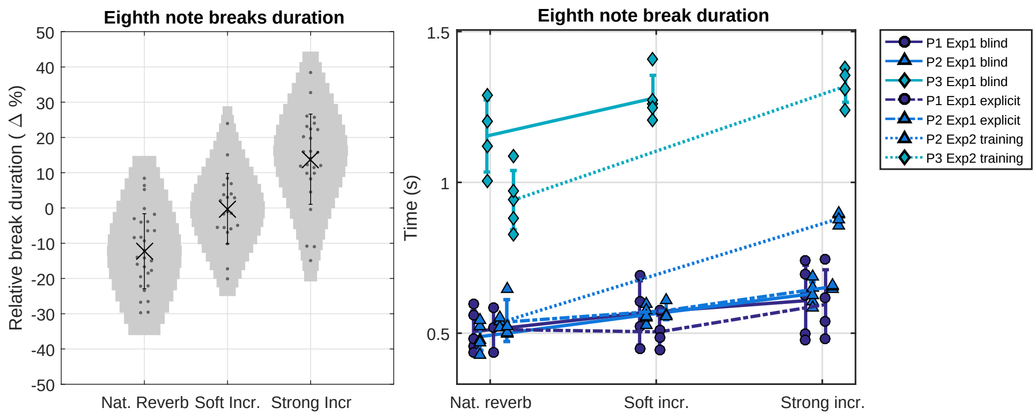


Fig. 6.27.: Duration of the eighth note breaks in *MendelssohnA* under different reverberation time.

- None of the players show a significant change in note duration when performing the piece *MendelssohnB*

- Player P4 tends to implement a more detached (*staccato*) articulation in all the performances when facing longer reverberation times. When performing the piece *Buxtehude1* those changes are not statistically significant.

- Player P3 does not present consistent note duration changes during the performance of the piece *Widor*.

- Both players performing *Bach2* (P2 and P5) present a tendency to reduce the duration of notes, although only the modifications implemented by P2 are statistically significant.

Breaks Duration

During musical breaks is when the decay of the sound is most audible, meaning that musicians and audience become more aware of the acoustic conditions of the room than during a continuous sound stream.

The start of the piece *MendelssohnA* is a succession of chords separated by eighth note breaks(see Fig. 6.26). To investigate the adaption strategies of musicians, the duration of the first five breaks has been extracted from the MIDI recordings and averaged.

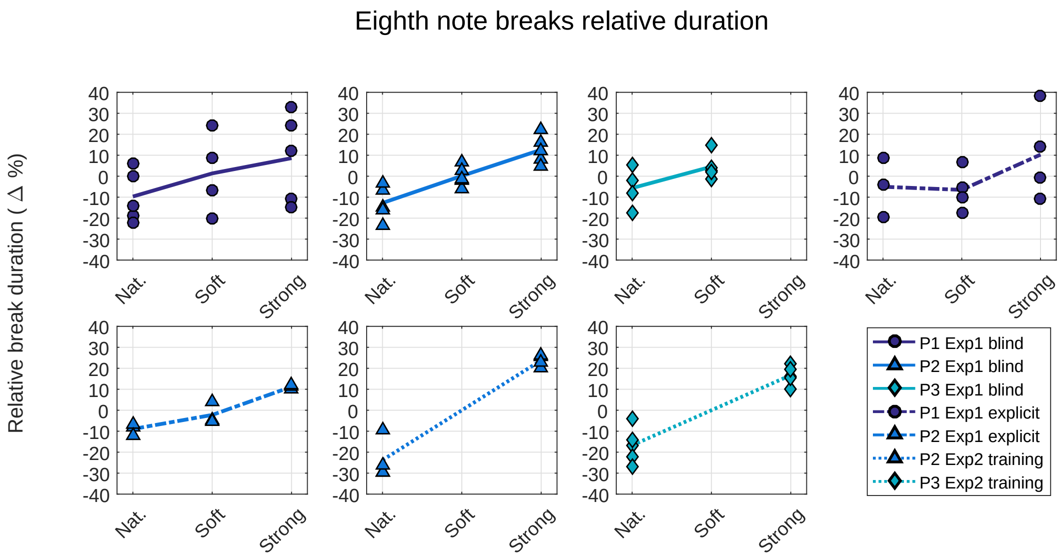


Fig. 6.28.: Relative duration of eighth note breaks in *MendelssohnA* under different reverberation time.

Results including all players that recorded this piece are presented in Fig 6.27. The left graph shows the relative break duration of all recorded players. It can be observed that there is a clear statistically significant tendency to extend the breaks in acoustic conditions with more reverberation. The average break duration in *Strong Increase* is approximately 25% higher than those recorded in *Natural reverb* conditions. The right graph presents the absolute value of the average break duration for every individual player. To analyze and compare players, the relative duration of breaks has been computed, and every player is presented separately in Fig. 6.28.

When analyzing the behavior of every player individually, the following can be concluded:

- All players tend to implement longer breaks when facing more reverberant conditions.
- Although player P1 follows this tendency, the changes are not statistically significant.
- Player P2 presents statistically significant changes in all experimental conditions (*blind*, *explicit* and *training*).
- Player P3 presents statistically significant changes in the condition *training*, and the changes are close to significance in the *blind* condition (p=0.1).
- The absolute duration of breaks in the performances of player P3 is considerably higher than the other players.
- The relative variation of breaks duration is highest for players P2 and P3 during the *training* experimental condition. The averaged relative differences in break duration when comparing *Natural Reverb* and *Strong Increase* exceed the 30%, and reaches close to 50% for player P2.

- Player P1 presents a high variance among average break duration, regardless of the acoustic conditions.

Piece	Player	Session	Condition	N	Tempo (p)	Note duration (p)	Break duration (p)
MendelssohnA	P1	Exp1	blind	14	0.46	0.78	0.30
MendelssohnA	P2	Exp1	blind	15	**0.00**	**0.00**	**0.00**
MendelssohnA	P3	Exp1	blind	9	0.09	0.91	0.10
MendelssohnA	P1	Exp1	explicit	11	<u>0.05</u>	0.15	0.33
MendelssohnA	P2	Exp1	explicit	9	**0.00**	**0.00**	**0.00**
MendelssohnA	P2	Exp2	training	8	<u>0.01</u>	0.56	**0.00**
MendelssohnA	P3	Exp2	training	10	**0.00**	<u>0.02</u>	**0.00**
MendelssohnB	P1	Exp1	blind	11	0.78	0.22	
MendelssohnB	P2	Exp1	blind	6	0.57	0.38	
MendelssohnB	P3	Exp1	blind	7	0.97	0.31	
Widor	P3	Exp1	blind	10	0.67	0.37	
Widor	P3	Exp2	training	4	0.18	0.57	
Bach	P4	Exp1	blind	8	0.38	**0.00**	
Bach	P4	Exp1	explicit	2			
Buxtehude1	P4	Exp1	blind	10	0.18	0.44	
Buxtehude2	P4	Exp1	blind	6	0.22	<u>0.01</u>	
Buxtehude2	P4	Exp1	explicit	2			
Bach2	P2	Exp2	training	4	0.57	<u>0.04</u>	
Bach2	P5	Exp2	training	6	0.29	0.25	

Tab. 6.18.: ANOVA p-values of the studied performance features in organ playing. Bold and underlined values refer to $p < 0.01$ and $p < 0.05$, respectively.

6.2.4 Interviews

During both sessions, interviews were conducted by W. Lachenmayr. The musicians were asked about their opinions regarding the acoustic conditions and the changes implemented in their performances. The interviews were conducted in german language and the relevant information was translated and collected later. The data of the interviews is presented here as a summary of the original conversations, and detailed data is presented in Tables 6.19, 6.20, 6.21, 6.22, 6.23, 6.24, and 6.25.

The main conclusions extracted from the interviews are:

- The adaption strategies and implemented changes reported by musicians are largely shared and appear to be common among all of them.

- The decrease of overall tempo is a typical approach when facing increased reverberation times. However, the nature of the piece is key in determining how much the tempo is decreased, if at all.

- If a piece presents sudden breaks where the reverberation is easily audible, the length of the notes preceding these breaks is largely affected, usually resulting in shorter notes if the reverberation is longer.

- A more detached or *staccato* articulation is a common approach when the reverberation time is higher.

- An increased amount of extra reverberation is usually preferred.

MendelssohnA	
Player	P1
Tempo	The player hears "when the room comes", and with more reverberation the breaks are extended. With drier acoustics they perform less *rubato*, the musical thinking is more string in temporal aspects.
Articulation	The notes are released earlier when the reverberation is longer, and the same phrases are played stronger or softer depending on the acoustics. Drier acoustics require a more *legato* articulation.
Acoustics	The *Soft Increase* condition is preferred. *Strong Increase* would require another registration with more 8-feet pipes, achieving a less transparent and more romantic sound.
Player	P2
Tempo	The playing style is changed in the frame of the available possibilities, and the tempo is slower. It is easy to hear the reverberation during the breaks, leading to longer breaks when the reverberation time is longer.
Articulation	The articulation in the lower voice is more *staccato*.
Acoustics	The *Soft Increase* condition is preferred. Longer reverberation contributes to a muddy sound, and some feedback is heard from the electro-acoustic enhancement.
Other	The different acoustic conditions can be easily identified, and after the two first chords the player reacts spontaneously to the acoustics.
Player	P3
Tempo	If there is no sufficient reverberation, the player needs to extend the notes before breaks, to compensate the fast sound decay.
Articulation	The reverberation creates an effect that has to be otherwise done by playing differently.
Acoustics	It is clearly better to have extra reverberation for this kind of romantic piece.

Tab. 6.19.: Musicians' verbal feedback related to the piece *MendelssohnA*

MendelssohnB	
Player	P1
Tempo	No tempo changes or intentional variations are mentioned.
Articulation	The left hand sounds stronger with increased reverb, and there are not noticeable changes in the right hand.
Acoustics	The *Strong Increase* condition is preferred, since it sounds nicer and "airy". The player described *Strong Increase* as a "sound carpet".
Player	P2
Tempo	No tempo changes or intentional variations are mentioned.
Articulation	In *Strong Increase* the left hand gets mixed and it is necessary to play more transparent, with smaller phrasing.
Acoustics	A difference is noticed when there is much reverberation (*Strong Increase*). The setting *Soft Increase* sounds more "organic" and is preferred, but the difference is subtle. The small differences could be due to the continuity of the sound during the piece, without long breaks.
Player	P3
Tempo	No tempo changes or intentional variations are mentioned.
Articulation	There are no important intentional variations in articulation, only a possible small effect on the melody voice, without mentioning further details.
Acoustics	Due to the nature of the piece there are not noticeable acoustic differences between conditions. When listening to their own MIDI recording the player preferred *Soft Increase*.

Tab. 6.20.: Musicians' verbal feedback related to the piece *MendelssohnB*

Widor	
Player	P3
Tempo	The tempo is not affected in this piece.
Articulation	The chords played by the left hand are more relaxed with increased reverberation. Without extra reverberation the first beat of every bar needs to be emphasized. The right hand (fast *staccato* notes) is easier to play with only natural reverberation.
Acoustics	Overall, since this piece is meant to be *staccato*, it is easier to play with only natural reverberation.

Tab. 6.21.: Musicians' verbal feedback related to the piece *Widor*

Bach	
Player	P4
Tempo	The tempo at beginning of the fugue feels better and steadier with increased reverberation.
Articulation	The articulation of the notes becomes shorter and more *staccato* with extended reverberation, including the pedals. The player deliberately playes longer notes in drier conditions, to compensate the effect of the missing reverberation.
Acoustics	Playing with longer reverberation is generally easier.

Tab. 6.22.: Musicians' verbal feedback related to the piece *Bach*

Buxtehude1	
Player	P4
Tempo	No tempo changes or intentional variations are mentioned.
Articulation	No intentional articulation changes are mentioned.
Acoustics	The difference between natural acoustic and enhanced reverberation is subtle, and the it is described as "something that is missing" when there is no enhancement.

Tab. 6.23.: Musicians' verbal feedback related to the piece *Buxtehude1*

Buxtehude2	
Player	P4
Tempo	The *ritardando* variations are stronger with increased reverberation.
Articulation	The overall articulation is more *staccato* with extended reverberation, specially on the fast notes of the right hand.
Acoustics	Specific comments about acoustic preference have not been given.

Tab. 6.24.: Musicians' verbal feedback related to the piece *Buxtehude2*

Bach2	
Player	P2
Tempo	No tempo changes or intentional variations are mentioned.
Articulation	The articulation is more "open" with extended reverberation.
Player	P5
Tempo	No tempo changes or intentional variations are mentioned.
Articulation	The articulation is different depending on the acoustics.
Acoustics	The task is easier with increased reverberation.

Tab. 6.25.: Musicians' verbal feedback related to the piece *Bach2*

6.2.5 Discussion

The results obtained from the analysis of the recordings suggest that organ players share adjustment strategies, which are mostly based on the decrease of *tempo* when performing in more reverberant conditions. However, when comparing the results of several pieces it can be observed that while the behavior is shared among multiple players, the musical character of the piece determines the importance of these adjustments. The *tempo* of *MendelssohnA*, a piece with full registration and a series of eighth note breaks after loud chords, is severely reduced and the duration of these breaks is increased. Contrarily, the next movement of the same piece, *MendelssohnB*, does not suffer any adjustment at all. In this case, *MendelssohnB* features quieter dynamics, a much softer registration and the absence of breaks, resulting into a continuous stream of sound. Thus, musicians have the opportunity to experience the characteristics of the reverberation during the breaks of *MendelssohnA*, but the acoustics blend together with the direct sound of the organ in *MendelssohnB*, difficulting the perception of the acoustic conditions. In addition, musicians mentioned that they use the acoustics as a musical resource, allowing the reverberation to fill the silence between consecutive notes, or waiting until a certain decay before starting the next chord. Thus, it can be expected that although players share adaption strategies, these will differ depending on the piece.

6.3 Conclusion

This chapter presented two approaches to the systematic study of performance adjustments of solo performers due to room acoustic conditions. While other studies were previously completed, the extent and research design of the present studies differ greatly from past work, thus contributing significantly to previous knowledge by generating new insights on the topic.

To the knowledge of the author, no previous formal studies on organ performance are available. However, it is worth doing a comparison of results between the presented experiments and previous studies featuring piano playing. While the role of the feet differs greatly in the playing technique of piano and organ, both instruments share important playing characteristics: both are key instruments, polyphonic and with a similar use of the hands. The most clear findings of the present studies refer to the reduction of *tempo* and increase of break duration in more reverberant conditions. Bolzinger *et al.* concluded that piano players tend to decrease the playing level [Bol+94] and overall *tempo* [BR92] in more reverberant rooms. Organ players are not able to dynamically modify the playing level during their performance to a great extent, given that they depend on the used registration. However, the behavior regarding *tempo* seems to be similar in organ and piano players. Moreover, Kawai *et al.* also stated that piano players tend to decrease the sustain pedal time in more reverberant conditions, which ultimately leads to longer musical breaks, similar to what organ players do when extend the duration of breaks in order to use reverberation as an aesthetic resource.

Two trumpet players participated in formal performance investigations with multiple instruments in the past [SK15]. However, with only two players and only two different pieces recorded by each player it was not possible to generalize the results, and the study concluded that although performance adjustments are systematic, they strongly depend on each player and instrument. The results presented in the present work allowed the generalization of at least two performance adjustments among trumpet players: the reduction of playing level and timbre in rooms with longer and stronger reverberation. Moreover, at least four of the studied players adjust significantly the played dynamics depending on the acoustic conditions. By having multiple pieces recorded by each player, with different musical characteristic, it has been possible to reduce the dependency of the results from the played musical excerpt. It has been shown as well that although it is generally not possible to classify trumpet players clearly by using all the implemented changes, it is indeed possible to observe generalized and shared behaviors among players when analyzing each performance dimension separately.

Performance adjustments of organ and trumpet players are indeed very contrasting, as are their playing techniques. However, among players of the same instrument it has been possible to identify generalized behaviors. This suggests that the playing technique and the family of the instruments play an important role on the nature of the adjustments that can be done. Thus, considering the present results, it can be assumed that the performance adjustments made by a violin and a violoncello player or those made by a trumpet and a trombone player, will be more similar than those made by a violin and a trumpet player, due to the similarities and differences of the mechanical processes involved in the playing technique of each instrument.

Finally, it should be noted that investigations regarding the impact of the musical characteristics of performed pieces should be further expanded. The present investigations demonstrate the importance of the pieces and present preliminary conclusions on potential musical characteristics that affect to the extent of the adjustments. In order to refine the characterization of performance adjustments these preliminary investigations should be expanded by studying a larger number of pieces, and the musical characteristics of the recorded excerpts must be included as a variable in the research design.

7 Perceptual evaluation of performance adjustments

As concluded from the experimental investigations presented in Chapter 6, and previously stated by various researchers [Uen+10; SK15], musicians implement systematic adjustments during their musical performance in order to accommodate the room acoustical conditions. Although little research has been conducted in order to evaluate the perceptual relevance of these adjustments from a listener perspective, Ueno *et al.* were able to conclude that in the majority of the cases subtle (44% of the subjects) to clear (54% of the subjects) differences can be perceived when comparing violin recordings conducted under different acoustic conditions. However, it was not stated explicitly which musical features were responsible for those differences, and instead every listener was encouraged to provide free feedback. This chapter presents two perceptual studies where listeners evaluate a set of trumpet and organ recordings obtained as a result of the experimental sessions presented previously.

Given that organ and trumpet are substantially different in many aspects e.g. sound generation, playing technique, polyphony... - and players of those instruments seem to adjust different musical aspects, the listener perception to those adjustments has been evaluated in two separate studies, one regarding organ recordings and one regarding trumpet recordings. The evaluation of organ recordings was completed during a session of the public workshop "WFS Spielräume" held regularly at the Detmold University of Music. As a part of the workshop, the attendants were invited to evaluate the differences between organ recordings. The evaluation of trumpet recordings was done through a typical listening test. The following sections describe further the implementation and results of the tests.

7.1 Organ Performance

As described in Chapter 7, organ recordings were conducted with musicians in the Konzerthaus of the Detmold University of Music in order to evaluate the effects of room acoustic conditions on the performance aspects of musicians. Those organ recordings were stored in binaural audio and MIDI formats, allowing binaural reproduction of the recordings at the musician position, but also allowing the reproduction of the MIDI files using the concert organ at the Konzerthaus. Thus, using the same set-up used during the experiments with musicians to modify the room acoustic conditions of the hall, the real performances could be exactly reproduced at any point using the MIDI playing capabilities of the organ.

Among the several recordings that resulted from the playing experiments, two pieces were selected. Each of the pieces had two versions recorded by the same player in different acoustic conditions - natural acoustics and soft increase (see Section 6.2 for a description of the conditions). The selected pieces were *MendelssohnA* and *MendelssohnB*, due to their distinct musical character and the fact that multiple players performed them, showing the same trends in the adjustment of their interpretation when facing different acoustics. Regarding *MendelssohnA*, players tended to decrease considerably the overall *tempo* in conditions with longer reverberation time, increasing the duration of the eighth note breaks. On the contrary, players did not seem to consistently adjust their performance when playing *MendelssohnB*. A summary of these results is presented in Fig. 7.1 and 7.2, and Tab. 7.1 and 7.3, to allow an easier comparison with the feedback provided by participants of the listening test.

The test was conducted by the author as a part of the public workshop "WFS Spielräume", regularly held at the Konzerthaus of the Detmold University of Music. The topic of the session was the modification of room acoustics using electro-acoustic systems. Participants were introduced to the possibilities of real-time manipulation of room acoustics and presented with some listening examples. All the participants were familiar with classical music and live performance, and after discussing the potential effects of room acoustics on musicians, a listening session was conducted.

Each participant was given a comment form to provide feedback regarding perceived differences in the performance, differences in the acoustics of the hall, general comments, and preference. Then, they were presented with the two versions of the recordings, reproduced using the MIDI interface of the organ, together with the associated acoustics of the recording. Each version of the piece was played 3 times, alternating between the two samples (natural and soft increase).

7.1.1 Results

The results of the test are presented in Tab. 7.2 and 7.4, that contain the feedback provided by the listeners during the session. For clarity, the feedback of each piece is presented separatedly, and a synthesis of the most important findings is provided below.

MendelssohnA

This piece was consistently and significantly adjusted by all the recorded players, which tended to decrease the overall *tempo* by implementing longer breaks and, in some cases, longer notes. As can be seen in Fig. 7.1, the main differences between the two versions of the selected piece rely on the overall *tempo* and the duration of the first eight note breaks located at the beginning of the piece (see the score in Fig. 7.1). While version A (natural reverb) implements a faster *tempo* with shorter breaks, version B (soft increase) features considerably longer breaks and a lower *tempo* at the beginning, to finally converge with the *tempo* of A at the end. These differences are described and perceived by all the players (see Tab. 7.2), who tend to prefer the version B of the recording.

P2 MendelssohnA	A (Natural Reverb)	B (Soft Increase)
Overall Tempo (BPM)	85.9	81.6
Avg. note duration (s)	0.61	0.63
Median note duration (s)	0.55	0.54
Avg. break duration (s)	0.50	0.62

Tab. 7.1.: Average performance values of *P2 MendelssohnA*

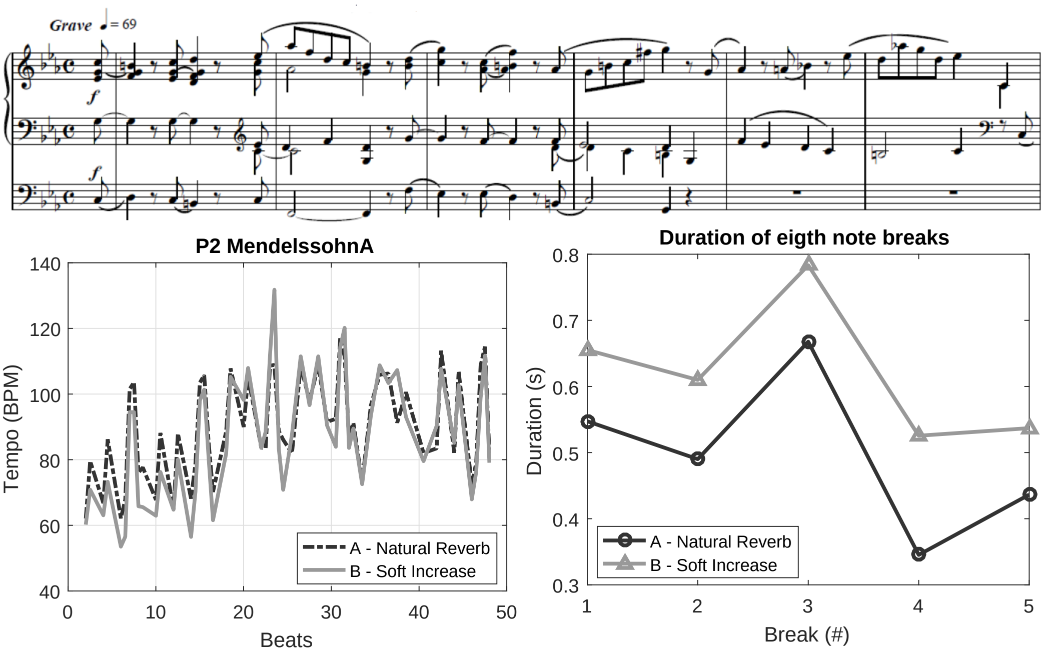


Fig. 7.1.: Score (top) and performance analysis (bottom) of *P2 MendelssohnA*

Subject	Comments	Differences in Performance	Differences in acoustics	Preferred piece
S1	Nicer sound in B. Both adapt, but greater effect with reverb	Strong agogic changes and longer breaks in B.	Extreme long reverb in B, moderate (or not extra reverb) in A.	B
S2		B starts slower	The sound has more time to build up in B.	B
S3		The first recording was faster and had shorter breaks (too short for my taste). The second player played longer notes.	The reverb in the second recording is much stronger than it is in the first one.	B
S4		A is faster, B is slower and fuller	Less sound in A	B

Tab. 7.2.: Listeners' feedback on *P2 MendelssohnA* compared recordings.

P2 MendelssohnB	A (Natural Reverb)	B (Soft Increase)
Overall Tempo (BPM)	76.9	76.6
Avg. note duration (s)	0.56	0.56
Median note duration (s)	0.39	0.39

Tab. 7.3.: Average performance values of *P2 MendelssohnB*

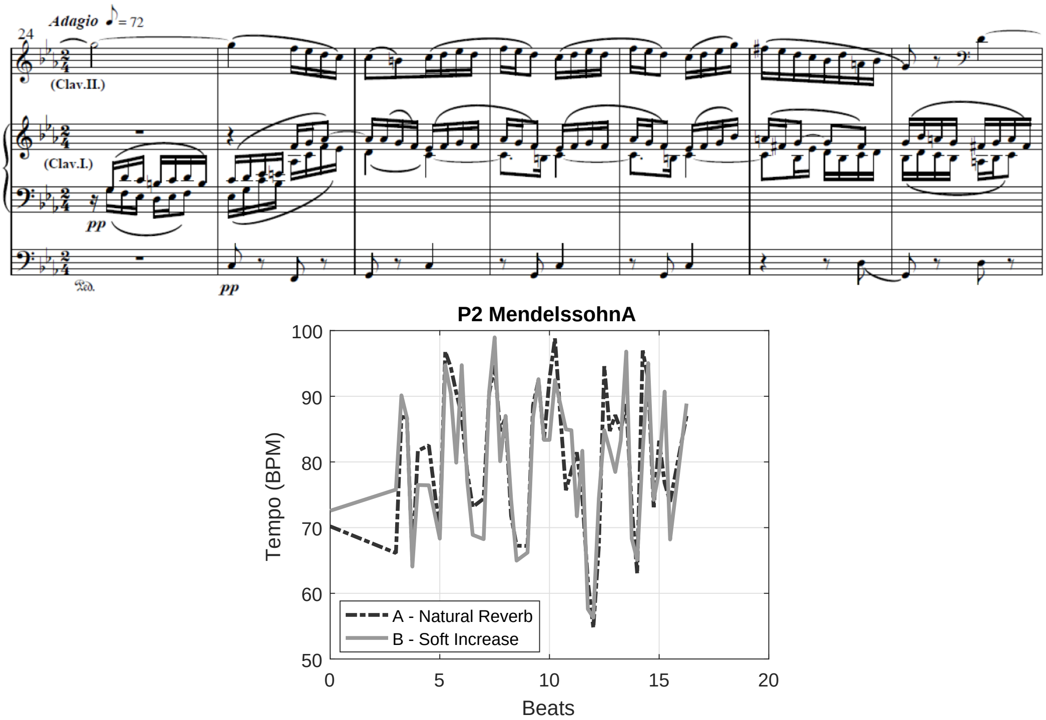


Fig. 7.2.: Score (top) and performance analysis (bottom) of *P2 MendelssohnB*

MendelssohnB

Contrarily to the previous piece, *MendelssohnB* was not systematically adjusted by the musicians. A possible explanation for this is the distinct musical character that these two pieces have. While *MendelssohnA* includes a series of strong chords at the beginning with full registration, *MendelssohnB* is much softer with more *legato* articulation and a more constant sound level, thus making it more difficult for musicians to perceive the length of reverberation. Accordingly, and although the judged recordings were different takes, the aspects reported by listeners refer to subtle differences between recordings and they do not show agreement. However, they all reported that the reverberation characteristics of the recordings were slightly different.

7.2 Trumpet Performance

During the playing experiments conducted with trumpet players, musical performances in four different acoustic conditions - *Dry*, *BS*, *DST*, and *KH* - were recorded. From the resulting dataset of anechoic recordings a reduced set of pieces was selected to implement formal listening tests, conducted using an on-line interface.

Subject	Differences in Performance	Differences in acoustics	Comments	Preferred piece
S1	Very small differences. Maybe more unbalanced note durations between the long and short notes in B.	Medium reverb in B, natural in A. The reverb is almost not audible, but changes the timbre	Artificial reverb adds a strange interference and sound coloration	A
S2	A has a little more rubato in the lower voices	In A: sustained noted a little bit less penetrant	Unbalanced: High voice too loud	A
S3	Both were similar, A was maybe a little bit softer	First recording has reverb but you can perceive it only by listening to the last note	Even if there are some differences because of the long notes, they are not easily perceived	-
S4	B has a fuller sound	B has a fuller sound		B

Tab. 7.4.: Listeners' feedback on *P2 MendelssohnB* compared recordings.

The goal of the test is double: to investigate whether small performance adjustments are perceived similarly by listeners, and to evaluate the relationship between the performance dimensions created by the MFA analysis in Chapter 6 and the perceptual impression of listeners regarding musical aspects.

7.2.1 Experiment Description

The experiment consisted on rating a number of musical aspects using a bipolar scale. The judged aspects are typical concepts commonly used in the musical vocabulary: Overall loudness, overall *tempo*, articulation, dynamic variations, sound color, *tempo* variability, and expressivity. However, while concepts such as overall *tempo* or loudness can be easily described and interpreted, other aspects such as dynamic variations or expressivity may differ significantly between listeners. For this reason, explicit descriptions of the aspects were not provided, instead, every bipolar scale was tagged at 7 points, similar to a Likert scale eg. for overall loudness the tags were: *much quieter, considerably quieter, slightly quieter, same loudness, slightly louder, considerably louder, much louder*. This encourages users to apply their subjective interpretation of the musical concepts, both in terms of quality and quantity, in an attempt to achieve a more natural approach, taking into account that music perception is substantially subjective, and thus should be reflected in the design of the test. It shall be noted, however, that users were encouraged to use the full range of the scales.

In order to center the range used by the participants, one of the recordings is used as a reference at the center of the scale and the three other recordings have to be rated. A screen capture of the test GUI is displayed in Fig. 7.3. Every judged parameter is organized in an independent axis, and each recording can be placed along the axis by moving the green marker.

Previous to the test, a personal survey has to be completed by every subject, including generic information - age, gender, known hearing problems, listening set-up information - and questions related to their musical background.

The test has been implemented using the Web Audio Evaluation Tool (WAET) [Jil+16] and uploaded for public access at an online server. In order to reach participants, the web address was distributed among musicians, recording engineers and audio scientists across a

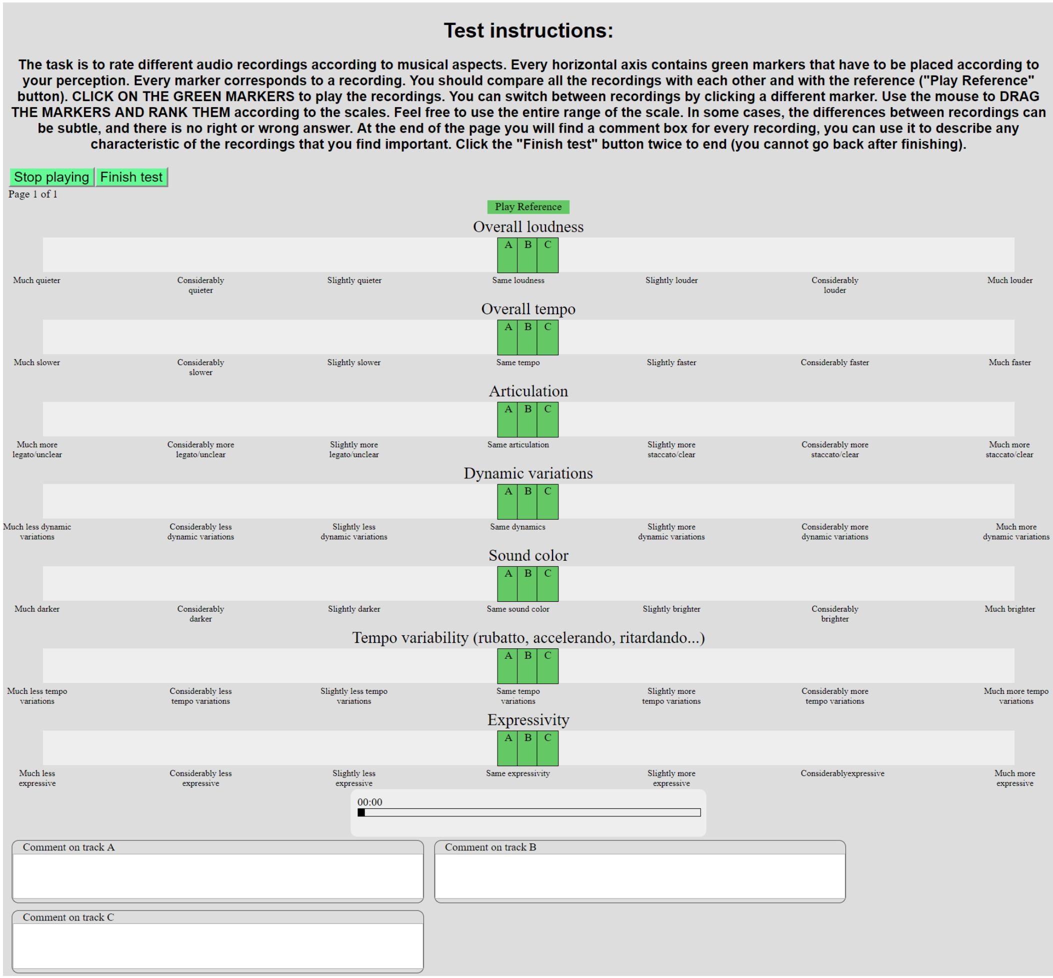


Fig. 7.3.: GUI of the online test

Piece, Movement	Composer	Bars	Comments
Slavonic Fantasy, Maestoso Sostenuto	Carl Höhne	1	*Cadenza*, various *fermatas*, free *tempo*, lyric.
Pictures at an Exhibition, Promenade	Modest Mussorgsky	1-4	Martial character with marked articulation.
Étude, 34	Théo Charlier	1-4	Various articulations, slow piece, many *fioriture*.

Tab. 7.5.: Piece excerpts used as stimuli in the listening test.

number of mailing lists, ensuring that potential participants had some experience with the topic at hand.

In order to evaluate the suitability of the test GUI, scales, rated parameters, and total duration, a pilot test was ran with the collaboration of 4 expert subjects familiar with critical listening and perceptual testing. The feedback of those subjects was used to refine the overall test procedure, GUI, used scales, and test instructions.

The listening test encouraged participants to express their feedback regarding the procedure and proposed tasks. Users with experience in music and critical listening tended to show a positive attitude towards the purpose of the test and the outcomes, while participants less familiar with classical music stated that the tasks might be too difficult for unexperienced users.

7.2.2 Stimuli

Among the approximately 360 recordings that were held during the experiments described in Section 6.1, a small subset was selected for the evaluation of listener perception to musical adjustments under different acoustics. The selection of the recordings was done considering the values of the MFA performance dimensions, assuming that higher differences in these values are perceptually more relevant than smaller differences. In addition, in order to evaluate a meaningful set of recordings, the evaluated excerpts have a distinct musical character. A summary of the pieces and their main characteristics is presented in Tab. 7.5.

The final set selected for evaluated consisted of 4 recorded versions of 3 pieces (one in each room, 12 recordings in total). The selected stimuli correspond to recordings of three different players, and all the versions of the same piece were recorded by the same player during the same session. It is worth noting that since the recordings were effectuated using a directional microphone close to the trumpet bell all of them exhibit anechoic conditions, and the differences between recordings are exclusively due to the playing adjustments of the musicians.

7.2.3 Participants

The test was completed by 24 participants with an average age of 31.2 years and a standard deviation of 11.7 years. Most of the participants were male and used headphones for the reproduction of the audio excerpts. A summary of the participants' data regarding previous

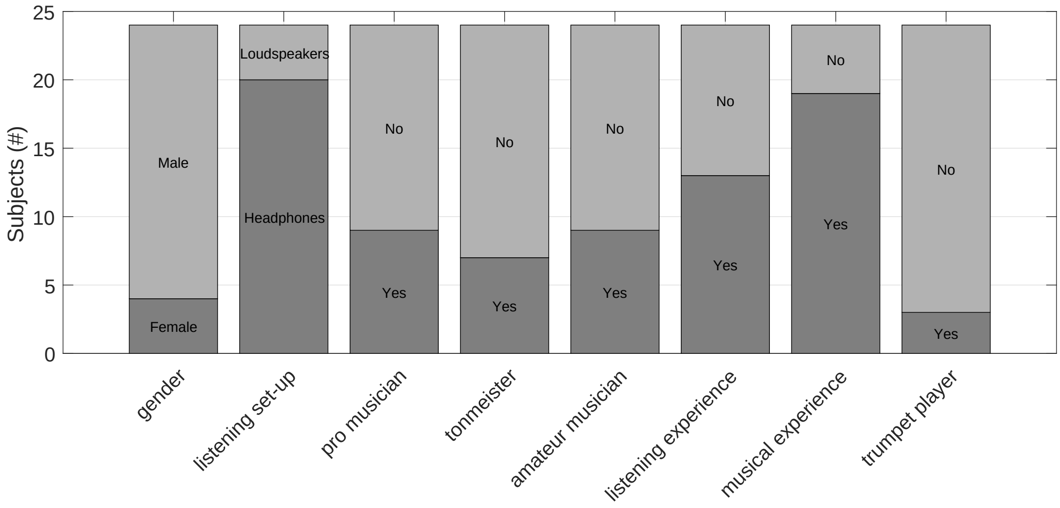


Fig. 7.4.: Data regarding the participants of the listening test.

experience with critical and musical listening is included in Fig. 7.4. As extracted from the data, a vast majority of the participants (19 out of 24) had previous musical experience.

7.2.4 Results

The results of the listening test are presented in Fig. 7.5 to Fig. 7.9. The graphs display the subjective ratings of each musical characteristic as a function of its respective MFA performance dimension. In the graphs, horizontal lines with asterisks denote statistical significant differences in the mean subjective rating of an evaluated feature (* refers $p < 0.05$ and ** refers to $p < 0.01$). Additionally, a linear fitting has been applied to the subjective data, and the adjusted R^2 is reported in each evaluated feature. Due to the small number of data points and given that a reference stimulus was used in the test to set the center of the rating axis, the presented data is not normalized, preserving inter-subject variability in the used range. The vertical boxes represent the 25th and 75th percentile, and the whiskers extend to the most extreme data points, leaving out the outliers.

The results serve a double goal: evaluating the perceptual impact of interpretative differences in recordings and the assessment of the goodness of fit of the MFA dimensions and the subjective perception of musical characteristics. In regards to the latter, Tab. 7.6 reports the Pearson correlation coefficients between the MFA dimensions ant the subjective ratings.

Piece	Dim. 1 Loudness	Dim. 1 Timbre	Dim. 2 Dynamics	Dim. 3 *Tempo*	Dim. 4: *Tempo* var.
T2 Hoehne1	**0.59**	**0.44**	0.13	-0.02	0.15
T5 Mussorgsky	0.49	0.43	0.61	-0.52	0.35
T9 Charlier34	**0.72**	**0.61**	0.46	-0.042	0.42

Tab. 7.6.: Pearson correlation coefficient between the MFA dimensions and the subjective ratings of their associated musical features. Significant correlations (p<0.01) with Holm–Bonferroni correction are expressed in bold font.

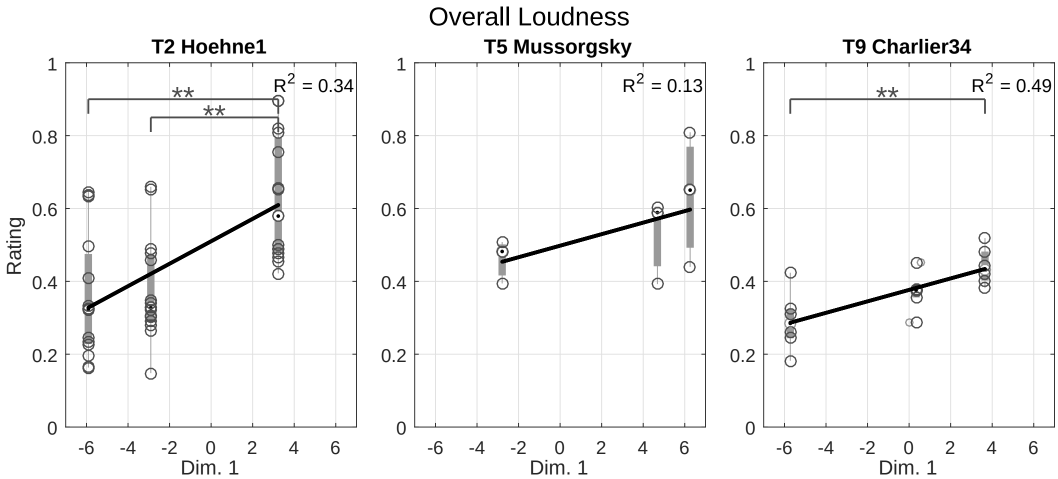


Fig. 7.5.: Perceived overall loudness of the evaluated excerpts. Brackets connecting observations denote statistically significant differences (* denotes $p < 0.05$; ** denotes $p < 0.01$).

Overall Loudness

The results (see Fig. 7.5) suggest that subtle interpretative adjustments regarding loudness are generally perceivable. Some samples of the stimuli *T2 Hoehne1* and *T9 Charlier34* present statistically significant differences in average perceived overall loudness. In addition, moderate to strong correlations, ranging between 0.49 and 0.72, are found between the perceived loudness and the MFA dimension 1, suggesting MFA is an appropriate method to construct perceptually relevant dimensions while reducing the dimensionality of the data.

Sound Color

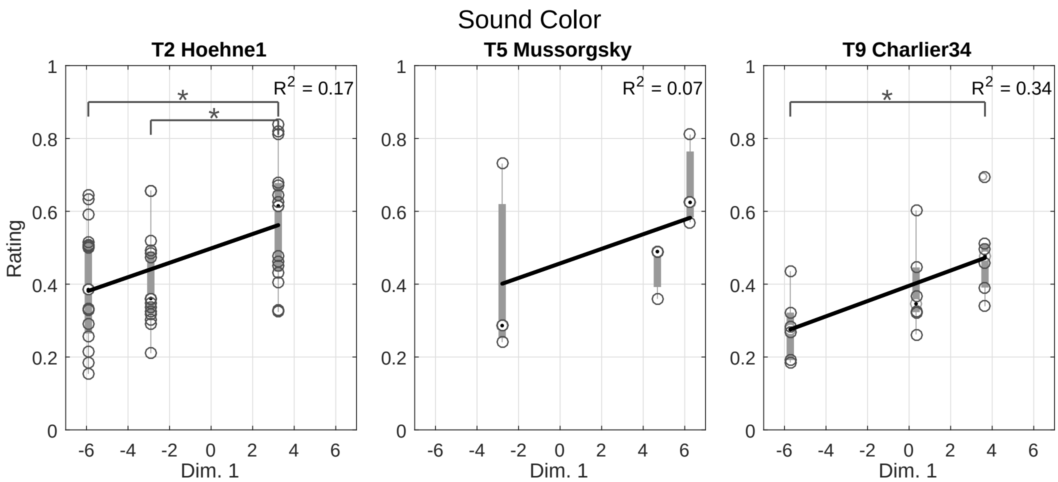


Fig. 7.6.: Perceived sound color of the evaluated excerpts. Brackets connecting observations denote statistically significant differences (* denotes $p < 0.05$; ** denotes $p < 0.01$).

Sound color refers to timbral aspects of the stimuli, and the usual qualitative terms at the end of a scale evaluating sound color scale range from *dark* to *bright*. The harmonic content of the spectrum of many instruments, including trumpet, increases proportionally to the sound level [Luc75]. The results regarding sound color (see Fig. 7.6) reaffirm that this physical phenomenon is consistently perceived by human listeners, as the subjective ratings regarding sound color present moderate and strong statistically significant correlations with MFA dimension 1. Thus, one dimension combining energy and timbral parameters is sufficient to explain level and timbral changes in trumpet performance. In addition, stimuli *T2 Hoehne1* and *T9 Charlier* present statistically significant perceivable differences in sound color due to interpretative variations.

Dynamics

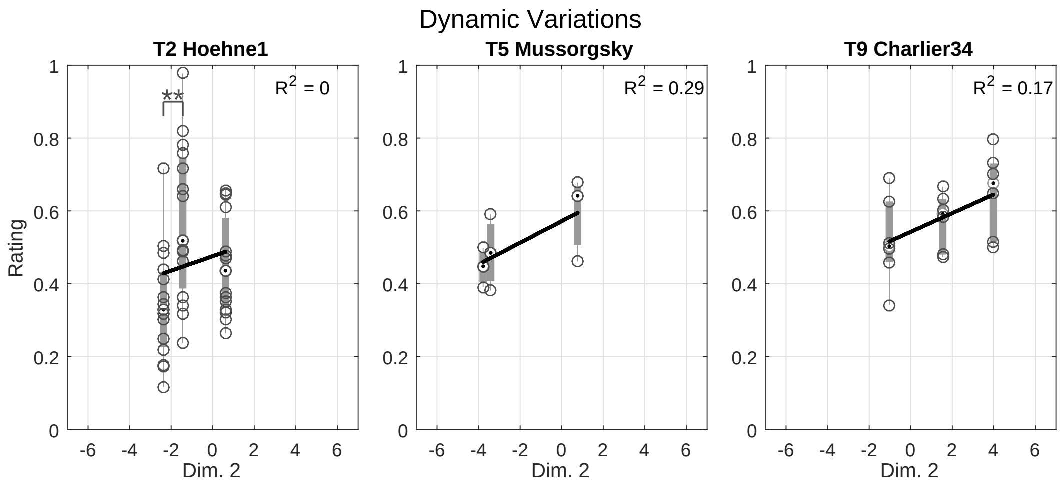


Fig. 7.7.: Perceived dynamic variations of the evaluated excerpts. Brackets connecting observations denote statistically significant differences (* denotes $p < 0.05$; ** denotes $p < 0.01$).

The perceptual ratings of dynamic variations (see Fig. 7.7) present moderate positive correlation with the values of MFA dimension 2 (see values in Tab. 7.6). In addition, compared recordings of *T2 Hoehne1* present statistical significant differences in perceived dynamic variations, suggesting that in some cases the interpretative adjustments are perceivable. However, the adjusted R^2 values of the linear fitting between the MFA dimension 2 and the perceptual ratings are generally lower to those relating overall loudness and sound color with dimension 1. A possible explanation for this is that while overall loudness and sound color are basic musical features presenting a more homogeneous definition among musicians and listeners, dynamic variations are intrinsically more complex. Whereas rating overall loudness and sound color respond to the evaluation of an average value of a specific musical parameter, the perception of dynamic variation can have a more subjective interpretation, with some subjects focusing on the dynamic range of the performance i.e. comparing *pianissimo* (minimum) and *fortissimo* (maximum) levels, other subjects focusing on the rate or frequency of dynamic differences, or a combination of range and frequency. Thus, it is possible that the musical interpretation of dynamic variations can not be reduced to the use of a single dimension, and it should be then characterized using a multidimensional space.

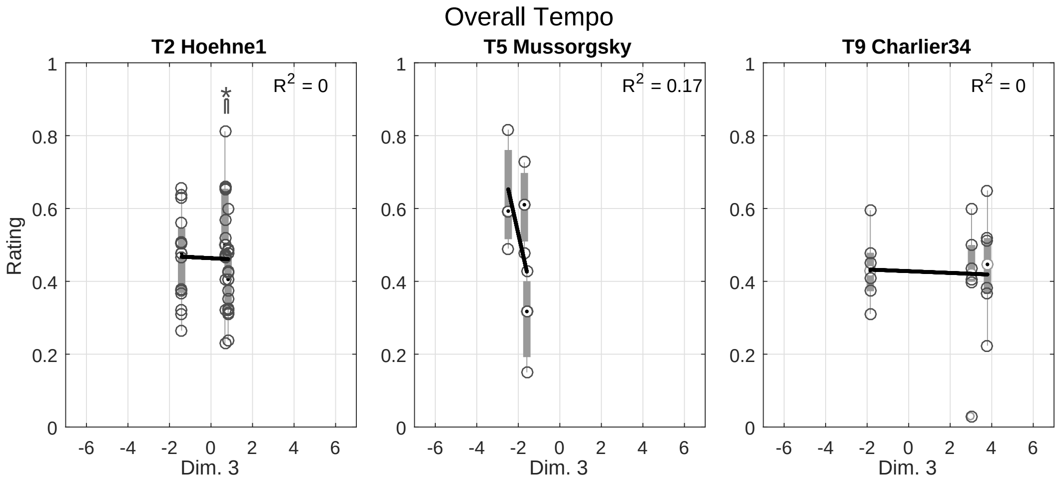


Fig. 7.8.: Perceived overall *tempo* of the evaluated excerpts. Brackets connecting observations denote statistically significant differences (* denotes $p < 0.05$; ** denotes $p < 0.01$).

Overall *tempo*

The results (see Fig. 7.8) suggest that interpretative differences regarding overall *tempo* are in general perceptually not relevant. In addition, the relationship between MFA dimension 3 and perceived overall *tempo* does not seem clear at first glance. Only the ratings of the stimulus *T5 Mussorgsky* lead to a moderate non statistically significant correlation between perceived overall *tempo* and MFA dimension 3. Since the audio features contributing to the construction of MFA dimension 3 are mostly total length of the excerpt and average *tempo* (in beats per minute), one would think that the relationship between those and the perceived *tempo* should be clear. Then, there are two proposed hypotheses for these results: either the perception of overall *tempo* responds to a more complex process and can not be expressed as a single value, or the *tempo* differences between stimuli are too small to lead to meaningful correlations.

Tempo variability

Similar to the case of dynamic variations, the *tempo* variability is subject to have a wide range of interpretations. Thus, a listener could focus on the range of temporal deviations from the average tempo, or instead focus on the frequency of those *tempo* adjustments (or a combination of both). The results of the listening test (see Fig. 7.9) suggest that there is no generalized agreement on the perceptual impact of *tempo* variations due to different acoustics. However, moderate (non statistically significant) correlations are present between the perceptual ratings and the MFA dimension 4. This suggests that, as in the case of dynamic variations, a multidimensional space could be better suited to describe the perceptual aspects of *tempo* variations.

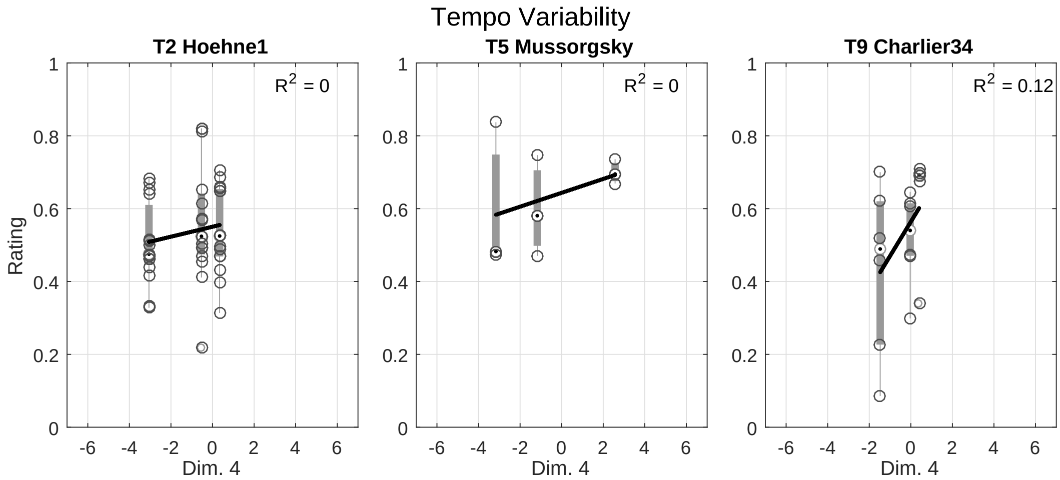


Fig. 7.9.: Perceived *tempo* variations of the evaluated excerpts. Brackets connecting observations denote statistically significant differences (* denotes $p < 0.05$; ** denotes $p < 0.01$).

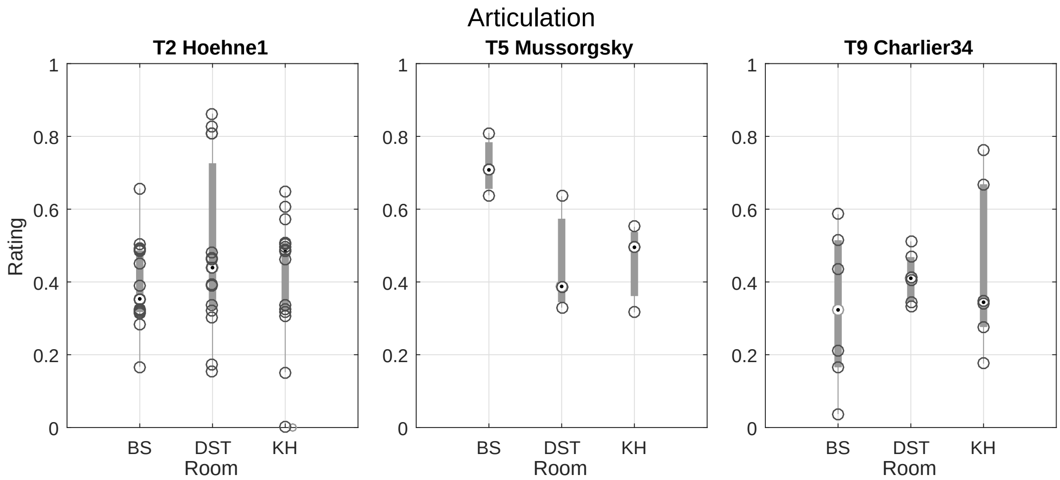


Fig. 7.10.: Perceived articulation of the evaluated excerpts. Brackets connecting observations denote statistically significant differences (* denotes $p < 0.05$; ** denotes $p < 0.01$).

Other features

The listening test analyzed as well the perceptual impression of articulation and expressivity. The results do not show a generalized perceptually relevant difference among the judged *stimuli*. Since these aspects have not been related previously to any of the constructed MFA dimensions, the results (see Fig. 7.10 and 7.11) are not mapped against the MFA values, and instead are labeled using the room name corresponding to the acoustic presented to the musicians during the recordings.

Neither articulation or expressivity seem to be generally perceived differently in the judged recordings, as no statistically significant differences are found.

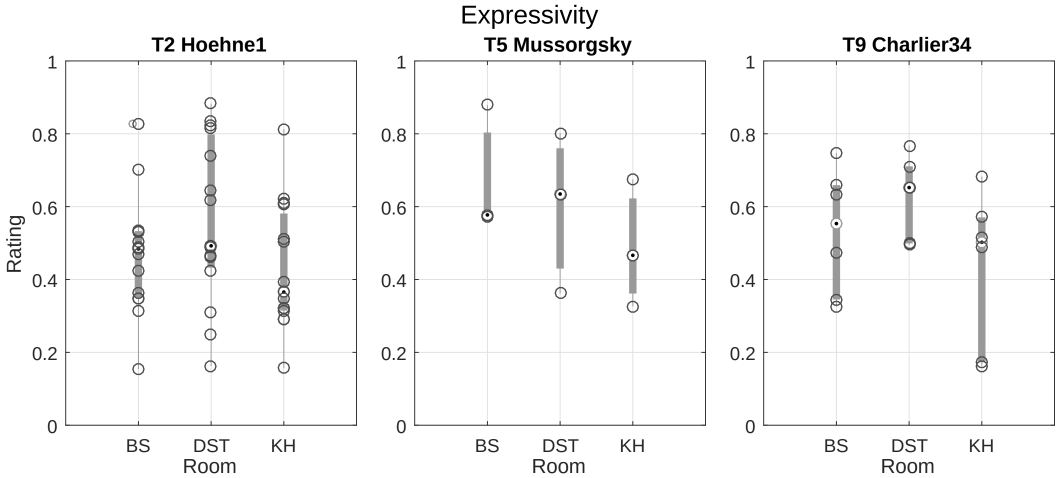


Fig. 7.11.: Perceived expressivity of the evaluated excerpts. Brackets connecting observations denote statistically significant differences (* denotes $p < 0.05$; ** denotes $p < 0.01$).

7.3 Discussion

As reported in Chapter 6, in the case of organ players, the main playing adjustments refer to *tempo* and articulation variations, while trumpet players mostly adjust the overall playing level, timbre (or sound color) and, to some extent, dynamics. Comparing these observations with the perceptual data collected during the experiments presented in the previous sections, one can conclude that interpretative adjustments implemented due to room acoustic variations are generally perceived by listeners, at least to some degree.

However, there are some performance aspects, or musical features, such as dynamic variations, *tempo* variations, or expressivity, that leave room for a personal interpretation. Thus, while trying to characterize these features with single unidimensional values nuances of this interpretation might be lost, and thus the subjective ratings of multiple listeners should be interpreted depending on the personal definition that every listener may have of each musical feature. This knowledge can be used for future tests, where the same procedure might be applied, but listeners should be asked to report about the personal definition that they apply in every musical feature. This would then allow the classification of different interpretations and a more thorough analysis of the test results.

Knowing that room acoustics have an effect on the emotional impact of orchestral music on listeners [PL16], future work could include the comparison between the presented results and the evaluation of musical features on the same stimuli including auralized acoustics from the listeners' side. This would allow a characterization of the impact of room acoustics on listeners' judgments of musical features.

8 Conclusion

This thesis represents an approach to the assessment of the effects of room acoustic conditions on live performances. Given that this is a highly interdisciplinary topic, the research work has been divided into four main blocks. The first part of the thesis presented the D3S, an auralization system that allows the resynthesis of room acoustic conditions in real-time. The second and third part of the work consist of experimental pilot studies evaluating the stage acoustic preferences and the study of performance adjustments using virtual acoustic environments. Finally, the perceptual aspects of performance adjustments from the listener perspective are assessed.

8.1 Summary of findings

The most important findings extracted from the results of the work are summarized below:

The usage of virtual acoustic and hybrid environments have proven to provide plausible acoustics and are appropriate methods to study performance. During the experimental studies, musicians consistently provided positive feedback regarding the acoustic qualities of the auralized spaces. The average grade of perceived realism of the resynthesized spaces from a survey evaluating the implemented system is 81/100. In addition, in spite of playing inside a loudspeaker set-up, with a microphone attached to the instrument, the set-up is widely considered not intrusive and does not disturb musicians while they play (average grade of intrusiveness – 17/100). Although virtual acoustic environments result in a lack of consistent auditory and visual feedback when performing, this is rapidly overcome by the fact that acoustics play a more important role during music performance. In addition, all the participating musicians stated that continuous access to such systems could contribute positively to training of musicians, as they are able to experience distinct acoustics in real-time without the necessity of physically moving to another rooms. Thus, this allows the exploration of acoustics and familiarization to different conditions in a way that is traditionally impossible. The implementation of the D3S represents a direct contribution to the research field of room acoustics and musical performance, opening the possibility of conducting systematic studies with musicians.

The study of stage acoustic preferences of solo trumpet players demonstrated that the performance context impacts significantly on the nature of acoustic conditions that musicians prefer. In this sense, while very dry acoustics usually lead to an increase of fatigue and are generally despised, they could be preferred under specific conditions. For instance, the study of instrument technique under dry acoustics allow musicians to concentrate on aspects of the performance that are usually masked or modified by reverberation, such as sound color or articulation. Contrarily, when judging concert conditions, usually longer reverberation times are preferred, allowing the musician to develop their musicality with increased freedom. A sufficient amount of early energy contributes positively to the comfort of musicians, although the absolute value is not easily quantifiable due to the calibration and miking process used in virtual acoustic environments. Preliminary results suggest that the direction of incidence of early reflections do not significantly affect the perceived stage support, provided that there is a sufficient amount of early energy. Although studies with a larger population are needed to confirm the last affirmation, to the best knowledge of the author, this is the first systematic study with musicians assessing directional properties of sound fields. In addition, these tests are among the first formal studies that evaluate stage acoustic conditions in auralized environments, and the presented methodology can be adopted in further investigations.

Musicians do consistently adjust their performance as a reaction to room acoustics. Two studies evaluating the performance adjustments of trumpet and organ players under different acoustic conditions have been completed, and systematic modifications of the performance have been identified. While past studies already concluded that musicians generally adjust their performance [Bol+94; Uen+10; Kat+15; SK15], the present work is the first featuring a larger population of trumpet and organ players. Hence, the results of these investigations constitute a direct increase of knowledge in the field.

The most generalized adjustments implemented by trumpet players consisted of a reduction of sound level and decrease of brightness in timbre during performance in more reverberant and energetic environments. Additionally, dynamic variations, *tempo* and temporal variations were systematically adjusted by a subset of musicians. This is extracted from an experiment consisting on the recording and analysis of live trumpet performances of 11 semi-professional musicians under different acoustic conditions, achieved by using a virtual acoustic environment. The analysis of the audio recordings consisted on the automatic extraction of 44 audio features that were linearly combined using MFA, reducing the dimensionality of the performance data. This resulted into 4 main performance dimensions: loudness & timbre, dynamic variations, *tempo*, and temporal variations.

As previous studies stated [SK15; Uen+10] it has been found that a clear classification of musicians into groups depending on their overall performance adjustment is generally not possible, due to the inherent complexity of a performance, individual behaviors and disparity of recorded excerpts. However, it has been found that **it is indeed possible to group and classify players using single performance dimensions (level, timbre, dynamics...) as a response variables to the predicting variable acoustics**. Additionally, in order to be able to identify further behavioral relationships between musicians, the musical piece could be used as a second predicting variable. However, achieving this would require all musicians to perform the same pieces, and it is out of the scope of this dissertation.

Organ players tend to modify temporal aspects of the performance depending on the acoustics of the room. This is concluded from a second playing experiment, featuring organ players in a hybrid environment (Detmold Konzerthaus using enhanced acoustics) and based on the analysis of MIDI recordings. The reverberation time of the room was used as a prediction variable, concluding that increase reverberation leads to a decrease of overall *tempo*. In addition, organ players tend to increase the duration of musical breaks when the reverberation is longer, using it as a musical resource. It is important to note that while these adjustments are generalized among all the studied players, the musical nature of the played pieces determines the extent of these adjustments, or the presence of them. For instance, soft pieces with *legato* articulation resulting into a continuous stream of sound are less prone to suffer performance adjustments than pieces with full registration and several musical breaks. The latter results in a higher excitation of the room, and in each of the breaks the musician can experiment the decay properties of the sound, thus impacting more prominently on the performance.

Listening tests concluded that listeners are generally able to perceive the performance adjustments made by musicians. In addition, the relationship between listener perceptual aspects of trumpet recordings and the analysis dimensions constructed using MFA has been analyzed, concluding that perception of sound color, loudness and dynamic variations correlate well with the constructed MFA dimensions. However, in some cases, the perception of dynamic and temporal variations could respond to multidimensional properties of the performance, meaning that absolute deviations of dynamics (quiet – loud) or *tempo* (slow – fast) and the frequency of those variations are both relevant for a listener, and the subjective interpretation of these musical features could depend on the musical context and have an individual interpretation for every listener. Although there have been previous efforts to model the perception and implement performance models of these musical features [BH10], at the moment of writing this dissertation there is no clear solution to modeling the perception of dynamics or *tempo* variations.

8.2 Further work

The auralization system implemented during this project provides plausible acoustics for trumpet players in real-time, by combining the use of a custom measurement set-up, room acoustic measurements, spatial analysis of impulse responses, real-time convolution and spatial reproduction of the convolved sound. In order to allow the use of this approach to other instruments, a number of elements of this process must be reevaluated. Directional sources must be used to excite the room as a real instrument would do, thus methods to generate them must be explored e.g. loudspeaker arrays or artificial excitation of real instruments. Additionally, the distances between source and receiver on stage should be modified as well to imitate stage set-ups of different instruments. This represents a problem of scalability, since several rounds of measurements with different sources and stage set-ups would be required in order to obtain SRIR suitable for a variety of instruments. Thus, the use of acoustic simulations instead of room acoustic measurements represents an alternative to conduct experiments with a larger set of instruments. Finally, the miking approach for instruments other than brass must be carefully studied, since the radiation properties are often more complex and change depending on the played note [PL10], thus multi-microphone approaches should be considered in wind instruments with tone holes or string instruments, among others.

The study of the relationship between room acoustics and musical performance has been approached from an empirical perspective with success. However, the amount of studied acoustic conditions and musical pieces was limited, and the expansion of these experimental conditions could contribute to a generalization of the results, by including more subjects, musical excerpts and room models to the presented studies.

The availability of a large database of recordings associated to different room acoustic conditions would represent an important contribution to the research of problems related to the perception of music, modeling of musical performances, automatic analysis of music or computer audition [WG04; EF17; McK05]. These problems usually benefit from the use of machine learning approaches, but large datasets are often required. The data collected during this work represents a first step towards the availability of this data.

Finally, next steps to expand the work presented in this document involves the study of musical ensembles, in terms of stage acoustic preferences, performance adjustments, and communication strategies. The auralization procedure should include reproduction methods allowing the presence of multiple listeners e.g. WFS, Ambisonics or other approaches with larger listening areas – and the measurement process should be re-designed to capture properly the response of the room at multiple playing position. However, the methodology used for the analysis of MIDI and audio recordings, and the implementation of systematic experiments could be reused to conduct studies with ensembles.

To enable further research in these directions and allow the reproduction of the presented experiments, the data generated during this project has been made public in an online repository [Ame17]. This includes measurement data and relevant scripts to implement auralizations, as well as the dataset of trumpet and organ recordings.

A

Measured room acoustic parameters

Freq. (Hz)	**EDT (s)**	**T_{20} (s)**	**T_{30} (s)**	**C_{80} (dB)**	**G_{all} (dB)**	**G_{early} (dB)**	**G_{late} (dB)**	**L_{early} (dB)**	**L_{late} (dB)**	**T_S (s)**
32	0.44	0.22		11.19	11.52	9.58	7.87	9.07	7.09	0.04
40	0.68	1.35	1.37	5.02	12.06	10.72	7.21	10.33	6.29	0.07
50	0.41	0.71	0.64	10.42	9.49	8.31	4.94	7.61	3.26	0.04
63	0.94	0.81	0.76	5.60	7.04	4.89	4.73	3.19	2.95	0.05
80	0.47	0.53	0.91	10.43	6.31	4.56	3.83	2.69	1.50	0.03
100	0.37	1.31	1.17	10.16	7.25	6.08	3.53	4.85	0.99	0.04
125	0.87	1.29	1.29	7.36	5.80	3.07	4.44	0.12	2.50	0.04
155	0.43	1.47	1.43	9.45	5.48	3.99	3.07	1.77	0.11	0.04
200	0.91	1.69	1.67	3.52	5.96	3.85	4.02	1.53	1.82	0.07
250	1.03	1.35	1.49	5.56	4.80	2.85	3.22	-0.35	0.40	0.06
315	1.29	1.50	1.55	4.45	3.36	1.36	2.55	-4.37	-0.97	0.05
400	1.01	1.32	1.42	5.91	4.63	2.88	2.93	-0.29	-0.17	0.05
500	0.91	1.26	1.31	4.97	5.14	3.31	3.26	0.56	0.49	0.05
630	1.14	1.37	1.35	4.99	3.78	1.78	2.74	-2.97	-0.57	0.05
800	1.18	1.21	1.20	4.69	3.23	1.37	2.38	-4.33	-1.37	0.05
1000	1.18	1.21	1.16	4.33	3.44	1.67	2.41	-3.36	-1.29	0.05
1250	1.32	1.14	1.11	3.81	3.68	1.85	2.56	-2.76	-0.97	0.06
1600	1.25	1.11	1.11	1.66	4.99	2.98	3.38	-0.17	0.67	0.08
2000	1.38	1.10	1.09	2.66	3.08	1.36	2.22	-4.43	-1.81	0.06
2500	1.36	1.08	1.10	2.08	3.79	1.90	2.67	-2.69	-0.83	0.07
3150	1.34	1.06	1.08	0.46	4.46	2.19	3.31	-1.86	0.50	0.08
4000	1.24	1.02	1.02	-0.12	4.65	2.46	3.36	-1.32	0.63	0.08
5000	1.18	0.96	0.95	-0.10	5.83	3.61	4.18	0.66	1.69	0.08
6350	1.13	0.87	0.87	1.09	4.78	2.93	3.13	-0.25	0.13	0.07
8000	1.01	0.76	0.75	-1.15	8.46	6.00	6.10	4.68	4.79	0.09
10000	0.90	0.64	0.64	-0.97	12.37	10.17	8.99	9.69	8.30	0.09
12500	0.88	0.50	0.51	-0.48	13.65	11.74	9.57	11.43	9.00	0.08
16000	0.76	0.38	0.39	2.73	12.47	11.37	6.78	11.01	5.60	0.06
20000	0.65	0.29	0.30	5.38	12.26	11.64	5.03	11.28	2.86	0.05

Tab. A.1.: Room acoustic parameters of the room BS

Freq. (Hz)	EDT (s)	T20 (s)	T30 (s)	C80 (dB)	Gall (dB)	Gearly (dB)	Glate (dB)	Learly (dB)	Llate (dB)	TS (s)
32	1.26	1.84	1.84	4.10	12.72	11.59	7.26	10.50	6.11	0.08
40	0.52	1.43	1.28	7.83	9.65	9.20	2.81	7.72	-0.91	0.05
50	0.42	0.85	0.73	9.68	8.92	8.51	2.32	6.89	-1.67	0.04
63	0.52	0.95	0.88	8.98	6.38	5.94	1.51	3.08	-3.89	0.04
80	0.37	0.92	0.94	12.64	3.40	3.24	0.33	-2.30	-11.24	0.02
100	0.15	1.10	0.95	15.82	3.17	3.06	0.22	-4.74	-12.93	0.02
125	0.12	1.01	1.24	16.11	2.53	2.44	0.15	-7.93	-14.50	0.01
155	0.45	0.91	1.03	10.41	2.22	1.92	0.46	-7.19	-9.48	0.02
200	0.25	1.04	1.02	13.76	1.39	1.23	0.21	-9.44	-13.07	0.02
250	0.20	1.22	1.21	13.76	0.69	0.54	0.17	-10.58	-13.92	0.01
315	0.06	1.02	0.97	15.45	0.55	0.46	0.10	-12.31	-16.32	0.01
400	0.33	0.90	1.02	11.50	1.12	0.96	0.21	-8.51	-13.12	0.02
500	0.05	0.93	0.91	14.16	0.46	0.36	0.11	-12.53	-15.89	0.01
630	0.11	0.90	1.05	14.62	0.48	0.38	0.11	-11.53	-15.91	0.01
800	0.31	1.06	1.01	12.59	0.55	0.39	0.18	-11.59	-13.75	0.01
1000	0.52	1.05	1.00	11.46	0.66	0.45	0.23	-10.12	-12.70	0.02
1250	0.74	0.99	0.99	9.87	1.08	0.74	0.40	-8.28	-10.16	0.02
1600	0.96	0.94	0.95	6.34	3.41	2.73	1.21	-1.31	-5.03	0.04
2000	1.09	0.91	0.93	9.16	1.25	0.81	0.52	-8.20	-9.04	0.02
2500	0.76	0.95	0.95	9.92	1.26	0.94	0.40	-7.61	-10.25	0.02
3150	1.16	0.93	0.92	7.72	1.64	1.12	0.69	-6.33	-7.93	0.03
4000	1.35	0.85	0.87	4.27	3.19	2.14	1.65	-2.90	-3.63	0.05
5000	1.26	0.83	0.83	6.48	2.08	1.38	0.94	-4.86	-6.30	0.04
6350	1.17	0.76	0.76	7.53	1.91	1.34	0.77	-5.38	-7.30	0.03
8000	1.06	0.63	0.65	4.95	4.19	3.26	1.79	-0.29	-3.04	0.05
10000	0.85	0.53	0.54	4.57	7.34	6.46	3.04	4.45	-0.05	0.06
12500	0.64	0.46	0.44	6.35	9.94	9.39	3.66	7.74	0.68	0.05
16000	0.45	0.41	0.36	9.27	9.73	9.46	2.09	7.86	-2.55	0.04
20000	0.38	0.37	0.32	12.39	9.00	8.88	0.92	7.14	-6.89	0.03

Tab. A.2.: Room acoustic parameters of the room DST

Freq. (Hz)	EDT (s)	T20 (s)	T30 (s)	C80 (dB)	Gall (dB)	Gearly (dB)	Glate (dB)	Learly (dB)	Llate (dB)	TS (s)
32	0.47	2.16	2.57	9.14	11.21	10.84	3.17	9.58	-1.12	0.05
40	0.43	1.91	1.60	11.16	9.07	8.85	1.46	7.22	-5.01	0.04
50	0.41	1.69	2.23	10.92	8.84	8.60	1.53	7.02	-4.01	0.05
63	0.31	1.49	1.68	10.34	5.42	5.12	0.89	1.66	-6.56	0.03
80	0.21	1.26	1.59	13.05	3.52	3.37	0.32	-2.07	-11.37	0.02
100	0.18	1.50	1.68	12.47	3.59	3.37	0.47	-3.44	-9.46	0.02
125	0.38	1.58	1.76	11.23	3.38	3.14	0.48	-4.56	-9.32	0.03
155	0.46	1.49	1.68	11.76	2.01	1.78	0.34	-8.40	-10.91	0.02
200	0.50	1.26	1.40	9.94	1.64	1.37	0.36	-7.00	-10.68	0.02
250	0.28	1.14	1.47	13.64	0.90	0.74	0.19	-9.20	-13.47	0.02
315	0.23	1.25	1.29	14.01	0.75	0.60	0.17	-9.89	-14.02	0.01
400	0.47	1.05	1.35	11.91	0.92	0.68	0.27	-9.17	-11.89	0.02
500	0.65	1.34	1.44	11.62	0.90	0.68	0.26	-9.20	-12.18	0.02
630	0.88	1.34	1.50	10.26	0.83	0.56	0.31	-9.27	-11.25	0.02
800	1.19	1.39	1.48	9.73	0.82	0.51	0.34	-9.30	-10.87	0.02
1000	1.56	1.35	1.48	8.00	1.00	0.48	0.58	-10.25	-8.50	0.03
1250	1.54	1.48	1.54	7.61	1.21	0.62	0.68	-8.56	-7.75	0.03
1600	1.61	1.44	1.48	3.66	2.80	1.63	1.62	-4.25	-3.55	0.06
2000	1.66	1.47	1.47	7.86	1.08	0.53	0.62	-9.25	-8.22	0.03
2500	1.45	1.42	1.40	7.21	1.47	0.87	0.73	-7.25	-7.57	0.04
3150	1.66	1.37	1.36	6.35	1.54	0.85	0.83	-7.16	-6.82	0.04
4000	1.60	1.28	1.27	4.23	2.31	1.27	1.37	-5.15	-4.52	0.06
5000	1.58	1.18	1.16	5.15	1.90	1.03	1.10	-6.11	-5.62	0.05
6350	1.49	1.04	1.02	6.76	1.60	0.92	0.83	-6.60	-6.88	0.03
8000	1.11	0.84	0.83	4.41	3.55	2.55	1.71	-1.40	-3.47	0.05
10000	0.93	0.66	0.67	2.89	7.70	6.46	3.89	4.96	1.56	0.07
12500	0.69	0.54	0.52	5.65	7.03	6.39	2.30	4.83	-1.71	0.05
16000	0.54	0.44	0.40	7.74	6.21	5.84	1.31	4.09	-4.92	0.04
20000	0.39	0.36	0.34	11.98	6.71	6.58	0.57	4.86	-8.90	0.03

Tab. A.3.: Room acoustic parameters of the room KH

B Virtual Environment - Technical Data

B.1 Equipment

	Computer 1	Computer 2
Type	Laptop	Desktop
CPU Model	Intel Core i7-4700MQ	Intel Core i5-6500
CPU Frequency	2.4 GHz	3.2 GHz
CPU Threads	8	4
RAM Memory	8 GB	16 GB
OS	Windows 10 Home	Windows 10 Enterprise
OS bits	64	64
Max/MSP Version	7.3.1 (64 bits)	7.3.1 (64 bits)

Tab. B.1.: Computers used in the real-time engine.

Device	Model	Serial number
Audio Interface	RME Madiface XT	–
Directional microphone	Schoeps CCM 4V	–
AD/DA converter	SSL Alphalink SX	–
Studio monitor (13 units)	Neumann KH120 A	–
Measurement microphone (6 units)	NTi M2010	3982, 3984, 3985, 3986, 3988, 4013

Tab. B.2.: Measurement and reproduction devices.

Device/Software	Channel	Level	Comments
RME Madiface XT	Analog Input	0 dB	
Max/MSP SphereConvolver	Input	126	
Max/MSP SphereConvolver	Output (all channels)	90	
Neumann KH120 A	All channels	108 dB	Flat EQ, fine gain min level

Tab. B.3.: Configuration values of the audio chain.

C Trumpet players data

Player	Age	Semester	Years playing	Concerts Solo	Concerts Chamber	Concerts Orchestra	Performances in KH	Performances in BS	Performances in DST
T1	23	1M	17	< 5	> 20	> 20	1 - 5	5 - 10	1 - 5
T2	20	5B	14	> 20	10 - 15	> 20	1 - 5	1 - 5	1 - 5
T3	20	3B	13	> 20	> 20	> 20	10 - 15	15 - 20	5 - 10
T4	25	7B	15	> 20	> 20	> 20	10 - 15	15 - 20	10 - 15
T5	23	7B	11	> 20	> 20	> 20	5 - 10	0	10 - 15
T6	31	2M	25	> 20	> 20	> 20	1 - 5	15 - 20	10 - 15
T7	25	1M	15	> 20	> 20	5 - 10	0	0	0
T8	21	5B	13	> 20	> 20	> 20	1 - 5	1 - 5	1 - 5
T9	27	7B	10	15 - 20	> 20	> 20	1 - 5	10 - 15	1 - 5
T10	22	4B	13	5 - 10	5 - 10	> 20	1 - 5	5 - 10	1 - 5
T11	20	5B	9	> 20	> 20	15 - 20	5 - 10	> 20	1 - 5

Tab. C.1.: Musical background of the participants in trumpet playing experiments.

Player	Importance of acoustics in solo performance	Importance of acoustics in orchestral performance	Do you adjust your performance according to acoustic conditions?
T1	80	90	Yes
T2	70	100	Yes
T3	50	50	Yes
T4	100	100	Yes
T5	60	70	Yes
T6	100	100	Yes
T7	100	100	Yes
T8	60	70	Yes
T9	80	90	Yes
T10	30	50	Yes
T11	100	70	Yes
Avg.	75	81	
Std. dev.	24	20	

Tab. C.2.: Subjective importance of acoustics.

D Preference Matrices

Practice Technique

	BS	DST	Dry	KH	Sum
BS	0	1	2	3	6
DST	13	0	7	13	33
Dry	12	7	0	12	31
KH	11	1	2	0	14

Tab. D.1.: Preference matrix of *Practice Technique* condition.

Practice Concert Piece

	BS	DST	Dry	KH	Sum
BS	0	5	7	2	14
DST	7	0	7	6	20
Dry	5	5	0	3	13
KH	10	6	9	0	25

Tab. D.2.: Preference matrix of *Practice Concert* condition.

Concert

	BS	DST	Dry	KH	Sum
BS	0	8	14	7	29
DST	6	0	14	5	25
Dry	0	0	0	0	0
KH	7	9	14	0	30

Tab. D.3.: Preference matrix of *Concert* condition.

Easiness

	BS	DST	Dry	KH	Sum
BS	0	9	14	9	32
DST	5	0	12	7	24
Dry	0	2	0	1	3
KH	5	7	13	0	25

Tab. D.4.: Preference matrix of *Easiness* condition.

Quality

	BS	DST	Dry	KH	Sum
BS	0	5	13	10	28
DST	9	0	14	8	41
Dry	1	0	0	2	3
KH	4	6	12	0	22

Tab. D.5.: Preference matrix of *Quality* condition.

E Musical Pieces

E.1 Organ pieces

Abbreviation	Composer	Piece	Movement	Bars
Bach	J. S. Bach	Prelude and Fugue in G Major, BWV 550	Grave and Fugue	0-39
Bach2	J. S. Bach	Fantasia and Fugue in G Minor, BWV 542	Prelude	1-13
Buxtehude1	D. Buxtehude	Prelude in F# Minor, Bux WV 146	Vivace	29-41
Buxtehude2	D. Buxtehude	Prelude in F# Minor, Bux WV 146	Vivace	42-end
MendelssohnA	F. Mendelssohn	Sonata Op.65 No.2	Grave	1-8
MendelssohnB	F. Mendelssohn	Sonata Op.65 No.2	Adagio	1-6
Widor	C. M. Widor	Organ Symphony No.5, Op.42	Toccata	1-11

Tab. E.1.: Pieces recorded during experiments with organ players.

E.2 Trumpet pieces

Abbreviation	Composer	Piece	Movement	Bars
Haydn1	F. J. Haydn	Trumpet Concerto in Eb Major	Allegro	37-50
Haydn2	F. J. Haydn	Trumpet Concerto in Eb Major	Andante	9-16
Hoehne1	C. Höhne	Slavic Fantasy	Maestoso Sostenuto	9
Hoehne2	C. Höhne	Slavic Fantasy	Maestoso Sostenuto	9-11
Tomasi	H. Tomasi	Concerto for Trumpet and Orchestra	Vivace	1-3
Albinoni	T. Albinoni	Concerto for in Sib	Allegro	56-64
VivaldiA	A. Vivaldi	Concerto in Re Minor	Vivace	20-36
Boehme1	O. Böhme	Concert, Op.18	Allegro moderato	8-19
Boehme2	O. Böhme	Concert, Op.18	Adagio religioso	7-15
Boehme3	O. Böhme	Concert, Op.18	Rondo	36-45
Hummel1	J. N. Hummel	Trumpet Concerto in E	Allegro con spirito	66-83
Charlier03	T. Charlier	36 Études transcendantes	No.3 "Intervalles"	1-24
Charlier2	T. Charlier	36 Études transcendantes	No.2 "Du style"	1-16
Bizet	G. Bizet	Carmen	Trumpet Call	-
Donizetti	G. Donizetti	Don Pasquale	Act 2, Preludio, Scene and Aria	5-12
Mussorgsky	M. Mussorgsky	Pictures at an exhibition	Promenade	1-4
Charlier16	T. Charlier	36 Études transcendantes	No.16 "Du staccato binaire"	1-38
Concone12	G. Concone	Lyrical studies	No.12, Allegretto brillante	1-48
Scelsi	G. Scelsi	Quatre pezzi	No.1	1-2
AllOfYou	C. Baker	Solo in "All of you"	-	1-30
DancoSamba	T. Bronner	Solo in "So Danco Samba"	-	1-31
SweetWayB	C. Baker	Solo in "In your own sweet way"	-	1-14
SoloInLove	B. Brookmeyer	Solo in "So in Love"	-	15-31
SweetWayH	T. Harrell	Solo in "In your own sweet way"	-	1-12
Piece2 VivaldiB	A. Vivaldi	Concerto in Re Minor	Vivace	20-36
Cavadini1	C. Cavadini	Sonatina: op. XVII	No. 1	-
Cavadini2	C. Cavadini	Sonatina: op. XVII	No. 2	-
Ropartz	G. Ropartz	Andante et Allegro	Allegro	5-25
Arabian	J. B. Arban	Arban's method for trumpet	Arabian song	-
Hansen	T. Hansen	Konzert Waltzer	-	10-27

Tab. E.2.: Pieces recorded during experiments with organ players.

Bibliography

[AB79] J. B. Allen and D. A. Berkley. “Image method for efficiently simulating small-room acoustics”. In: *The Journal of the Acoustical Society of America* 65.4 (1979), pp. 943–950 (cit. on p. 14).

[Abd+13] H. Abdi, L. J. Williams, and D. Valentin. “Multiple factor analysis: Principal component analysis for multitable and multiblock data sets”. In: *Wiley Interdisciplinary Reviews: Computational Statistics* 5.2 (2013), pp. 149–179. arXiv: `arXiv:1011.1669v3` (cit. on p. 70).

[Ame17] S. V. Amengual. “Investigations on the Influence of Acoustics on Live Music Performance using Virtual Acoustic Methods”. In: *https://doi.org/10.6084/m9.figshare.5454364.v2*. Figshare Repository, 2017 (cit. on p. 128).

[AP12] H. Arau-Puchades. “A Symphony Hall: L’Auditori Barcelona”. In: *Journal of Building Acoustics* 19.4 (2012), pp. 249–265 (cit. on p. 9).

[Bau61] B. B Bauer. “Stereophonic Earphones and Binaural Loudspeakers”. In: *Journal of the Audio Engineering Society* 9.2 (1961), pp. 148–151 (cit. on p. 14).

[Ber+93] A. J. Berkhout, D. de Vries, and P. Vogel. “Acoustic control by wave field synthesis”. In: *The Journal of the Acoustical Society of America* 93.5 (1993), pp. 2764–2778 (cit. on p. 14).

[Ber+97] A. J. Berkhout, Di. de Vries, and J. J. Sonke. “Array technology for acoustic wave field analysis in enclosures”. In: *The Journal of the Acoustical Society of America* 102.5 (1997), pp. 2757–2770 (cit. on p. 13).

[BH10] A. Berndt and T. Hähnel. “Modelling Musical Dynamics”. In: *Proceedings of Audio Mostly 2010: 5th Conference on Interaction with Sound - Sound and Design*. 2010, pp. 134–141 (cit. on p. 127).

[Bla69] J. Blauert. “Sound localization in the median plane”. In: *Acta Acustica united with Acustica* 22.4 (1969), pp. 205–213 (cit. on p. 28).

[Bla96] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1996, p. 508 (cit. on pp. 9, 15, 28).

[Boe01] P. Boersma. “Praat, a system for doing phonetics by computer”. In: *Glot International* 5.9/10 (2001), pp. 341–347 (cit. on p. 19).

[Bog+13] D. Bogdanov, N. Wack, E. Gómez, et al. “Essentia: An audio analysis library for music information retrieval”. In: *Proceedings of the International Conference on Music Information Retrieval (ISMIR)* (2013), pp. 493–498 (cit. on p. 19).

[Bol+94] S. Bolzinger, O. Warusfel, and E. Kahle. “A study of the influence of room acoustics on piano performance”. In: *Journal de Physique IV* 4.C5 (1994), pp. 617–620 (cit. on pp. 18, 20, 21, 108, 126).

[Bor84] J. Borish. “Extension of the image model to arbitrary polyhedra”. In: *The Journal of the Acoustical Society of America* 75.6 (1984), p. 1827 (cit. on p. 14).

[Bot95] B. Botteldooren. “Finite-difference time-domain simulation of low-frequency room acoustic problems”. In: *The Journal of the Acoustical Society of America* 98.6 (1995), pp. 3302–3308 (cit. on p. 14).

[BR15] J. Blauert and A. Raake. “Can Current Room-acoustics Indices Specify the Quality of Aural Experience in Concert Halls?” In: *Psychomusicology: Music, Mind, and Brain* 25 (2015), pp. 253–255 (cit. on p. 9).

[BR92] S. Bolzinger and J. C. Risset. “A Preliminary Study on the Influence of Room Acoustics on Piano Performance”. In: *Journal de Physique IV* 2.C1 (1992), pp. 93–96 (cit. on pp. 20, 21, 108).

[Bre+12] J. S. Brereton, D. T. Murphy, and D. M. Howard. “The Virtual Singing Studio : A loudspeaker-based room acoustics simulation for real- time musical performance The Virtual Singing Studio”. In: *Proceedings of the Joint Baltic-Nordic Acoustics Meeting*. 2012, p. 8 (cit. on p. 16).

[Bre14] J. S. Brereton. “Singing in Space(s): Singing performance in real and virtual acoustic environments — Singers’ evaluation, performance analysis and listeners’ perception .” PhD thesis. 2014, p. 312 (cit. on pp. 16, 19).

[Bur08] C. S. Burrus. *Convolution Algorithms*. 2008 (cit. on p. 29).

[Cab+10] D. Cabrera, R. McCarthy, J. Bassett, and D. Gilfilla. “The measurement of directional stage support in a drama theatre: quantifying the acoustic effect of a set”. In: *Journal of Building Acoustics* 17.1 (2010), pp. 1–14 (cit. on p. 9).

[Cab+11] D. Cabrera, D. Lee, M. Yadav, and W. L. Martens. “Decay Envelope Manipulation of Room Impulse Responses: Techniques for Auralization and Sonification”. In: *Proceedings of Acoustics 2011*. Gold Coast, 2011 (cit. on p. 24).

[Can+10] C. Cannam, C. Landone, and M. Sandler. “Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files”. In: *Proceedings of the ACM Multimedia 2010 International Conference* (2010), pp. 1467–1468 (cit. on p. 19).

[Chi+03] W. Chiang, S. T. Chen, and C. T. Huang. “Subjective Assessment of Stage Acoustics for Solo and Chamber Music Performances”. In: *Acta Acustica* 89.5 (2003), pp. 848–856 (cit. on p. 11).

[CO+15] J. J. Carabias-Orti, F. J. Rodriguez-Serrano, P. Vera-Candeas, N. Ruiz-Reyes, and F. J. Cañadas-Quesada. “An audio to score alignment framework using spectral factorization and dynamic time warping”. In: *ISMIR: proceedings of the International Conference of Music Information Retrieval*. Málaga, 2015, pp. 742–748 (cit. on pp. 19, 35).

[CT65] J. W. Cooley and J. W. Tukey. “An Algorithm for the Machine Calculation of Complex Fourier Series”. In: *Mathematics of Computation* 19.90 (1965), pp. 297–301 (cit. on p. 29).

[Dam09] J. J. Dammerud. “Stage Acoustics for Symphony Orchestras in Concert Halls”. PhD thesis. 2009, p. 210 (cit. on p. 69).

[Dam11] J. J. Dammerud. *Details on G versus ST as a measure of absolute acoustic levels*. 2011 (cit. on p. 66).

[DG02] S. Dixon and W. Goebl. “Pinpointing the beat: Tapping to expressive performances”. In: *Proceedings of the 7th International Conference on Music Perception and Cognition*. July. Sydney, 2002, p. 4 (cit. on p. 19).

[DI16] J. S. Downie and IMIRSEL. *MIREX 2016 Evaluation Results*. Tech. rep. University of Illinois at Urbana-Champaign, 2016 (cit. on p. 19).

[Dow03] J. S. Downie. "Music Information Retrieval". In: *Annual Review of Information Science and Technology* 37 (2003), pp. 295–340 (cit. on p. 18).

[Dow08] J. S. Downie. "The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research". In: *Acoustical Science and Technology* 29.4 (2008), pp. 247–255 (cit. on p. 19).

[DP08] G. Defrance and J. D. Polack. "Measuring the mixing time in auditoria". In: *Proceedings of Acoustics '08 Paris*. Paris, 2008, pp. 3871–3876 (cit. on p. 6).

[EF17] A. Elowsson and A. Friberg. "Predicting the perception of performed dynamics in music audio with ensemble learning". In: *The Journal of the Acoustical Society of America* 141.3 (2017), pp. 2224–2242 (cit. on p. 128).

[Ell03] D. Ellis. *Dynamic Time Warp (DTW) in Matlab*. 2003 (cit. on pp. 34, 35).

[ET04] T. Eerola and P. Toiviainen. "MIR in Matlab: The MIDI Toolbox". In: *Proceedings of the International Symposium on Music Information Retrieval Conference (ISMIR)*. 2004 (cit. on pp. 19, 30).

[Far00] A. Farina. "Simultaneous measurement of impulse response and distortion with a swept-sine technique". In: *108th AES Convention*. Paris, 2000 (cit. on pp. 23, 43).

[Far79a] K. Farrar. "Soundfield microphone - 2: Detailed functioning of control unit". In: *Wireless World* November 1 (1979), pp. 99–103 (cit. on p. 16).

[Far79b] K. Farrar. "Soundfield microphone: Design and development of microphone and control unit". In: *Wireless World* October 19 (1979), pp. 48–50 (cit. on p. 16).

[Fri+07] A. Friberg, E. Schoonderwaldt, and P. N. Juslin. "CUEX: An algorithm for automatic extraction of expressive tone parameters in music performance from acoustic signals". In: *Acta Acustica united with Acustica* 93 (2007), pp. 411–420 (cit. on pp. 31, 33, 35).

[Fri+11] A. Friberg, E. Schoonderwaldt, and A. Hedblad. "Perceptual ratings of musical parameters". In: *Gemessene Interpretation - Computergestützte Aufführungsanalyse im Kreuzverhör der Disziplinen*. Ed. by H. Loesch and S. Weinzierl. Schott (Klang und Begriff 4), 2011. Chap. Perceptual, pp. 237–253 (cit. on p. 18).

[Fri+14] A. Friberg, E. Schoonderwaldt, A. Hedblad, M. Fabiani, and A. Elowsson. "Using listener-based perceptual features as intermediate representations in music information retrieval". In: *Journal of the Acoustical Society of America* 136.4 (2014), pp. 1951–1963 (cit. on pp. 18, 19).

[Fry80] D. B. Fry. "The singer and the auditorium". In: *Journal of Sound and Vibration* 69.1 (1980), pp. 139–142 (cit. on p. 12).

[Gab99] A. Gabrielsson. "The Performance of Music". In: *The Psychology of Music*. Ed. by Diana Deutsch. 2nd Editio. Academic Press Series in Cognition and Perception, 1999, pp. 501–602 (cit. on p. 17).

[Gad89a] A. C. Gade. "Investigations of Musicians' Room Acoustic Conditions in Concert Halls. II: Field Experiments and Synthesis of Results". In: *Acustica* 69 (1989), pp. 249–262 (cit. on p. 11).

[Gad89b] A. C. Gade. "Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and Laboratory experiments". In: *Acustica* 69 (1989), pp. 193–203 (cit. on pp. 8, 11, 12, 15, 58, 63, 65).

[Gad92] A. C. Gade. "Practical aspects of room acoustic measurements on orchestra platforms". In: *Proceedings of the International Congress on Acoustics*. Beijing, 1992 (cit. on p. 8).

[Gar95] W. G. Gardner. "Efficient Convolution without Input-Output Delay". In: *Journal of the Audio Engineering Society* 43.3 (1995), pp. 127–136 (cit. on p. 29).

[GB03] W. Goebl and R. Bresin. “Measurement and reproduction accuracy of computer-controlled grand pianos.” In: *The Journal of the Acoustical Society of America* 114.4 (2003), pp. 2273–2283 (cit. on p. 19).

[Ger03] D. Gerhard. *Pitch Extraction and Fundamental Frequency: History and Current Techniques*. Tech. rep. University of Regina, 2003, pp. 1–22 (cit. on p. 19).

[Ger73] M. A. Gerzon. “Periphony: With-Height Sound Reproduction”. In: *Journal of the Audio Engineering Society* 21.1 (1973), pp. 2–8 (cit. on p. 14).

[GJ96] A. Gabrielsson and P. N. Juslin. “Emotional Expression in Music Performance: Between the Performer’s Intention and the Listener’s Experience”. In: *Psychology of Music* 24 (1996), pp. 68–91 (cit. on p. 17).

[Gut+13] A. Guthrie, S. Clapp, J. Braasch, and N. Xiang. “Using Ambisonics for Stage Acoustics Research”. In: *Proceedings of International Symposium on Room Acoustics*. 2013, pp. 1–10 (cit. on pp. 10, 60, 66).

[Had+13] A. Hadjakos, T. Großhauser, and W. Goebl. “Motion Analysis of Music Ensembles with the Kinect”. In: *Proceedings of the International Conference on New Interfaces for Musical Expression*. 2013, pp. 106–110 (cit. on p. 18).

[Had12] A. Hadjakos. “Pianist Motion Capture with the Kinect Depth Camera”. In: *Proceedings of the Sound and Music Computing Conference*. 2012, p. 8 (cit. on p. 18).

[Hat12] E. Hatlevik. *Are Musicians Affected by Room Acoustics in Rehearsal Rooms?* M.Sc. Thes. Norwegian University of Science and Technology (NTNU), 2012 (cit. on p. 21).

[HT12] A. Harker and P. A. Tremblay. “The HISSTools Impulse Response Toolbox: Convolution for the Masses”. In: *Proceedings of the International Computer Music Conference: Non-Cochlear Sound*. 2012, pp. 148–155 (cit. on pp. 29, 46).

[ISO09] International Organization for Standarization ISO. *ISO 3382-1:2009 - Acoustics. Measurement of room acoustic parameters. Part 1: Performance Spaces*. 2009 (cit. on pp. 6, 8, 9, 41, 66).

[ISO16] International Organization for Standarization ISO. *ISO 1996-1:2016 - Acoustics - Description, measurement and assessment of environmental noise*. 2016 (cit. on p. 47).

[JC91] J. M. Jot and A. Chaigne. “Digital Delay Networks for Designing Artificial Reverberators”. In: *90th Convention of the Audio Engineering Society*. Paris, 1991 (cit. on p. 16).

[Jeo+14] J. Y. Jeon, H. Lim, and Y. S. Kim. “Preferred acoustical conditions for musicians on stage in solo , duet and ensemble performances”. In: *Proceedings of the Forum Acusticum*. Krakow, 2014 (cit. on pp. 12, 58).

[Jil+16] N. Jillings, B. De Man, D. Moffat, J. D. Reiss, and R. Stables. “Web Audio Evaluation Tool : A framework for subjective assessment of audio”. In: *Proceedings of the Web Audio Conference*. Atlanta, 2016 (cit. on p. 115).

[Jus13a] P. N. Juslin. “From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions”. In: *Physics of Life Reviews* 10 (2013), pp. 235–266 (cit. on p. 17).

[Jus13b] P. N. Juslin. “What does music express? Basic emotions and beyond”. In: *Frontiers in Psychology* 4 (2013), pp. 1–14 (cit. on p. 1).

[Kah16] E. Kahle. “Acoustic feedback for performers on stage – return from experience”. In: *Proceedings of the International Symposium on Music and Room Acoustics*. La Plata, 2016 (cit. on pp. 9, 10).

[Kat+15] K. Kato, K. Ueno, and K. Kawai. “Effect of Room Acoustics on Musicians ’ Performance. Part II : Acoustic Analysis of the Variations in Performed Sound Signals”. In: *Acta Acustica united with Acustica* 101.4 (2015), pp. 743–759 (cit. on pp. 8, 19, 20, 35, 126).

[Kaw+13] K. Kawai, K. Harada, K. Ueno, T. Sakuma, and K. Kato. "Experiment on the Adjustment of Piano Performance To Suit Room Acoustics: Analysis of Performance Coded Into Midi Data". In: *Proceedings of the International Symposium on Room Acoustics*. Toronto, 2013 (cit. on pp. 18, 21).

[KL17] A. Kuusinen and T. Lokki. "Wheel of Concert Hall Acoustics". In: *Acta Acustica united with Acustica* 103.2 (2017), pp. 185–188 (cit. on pp. 10, 61).

[Kle+93] M. Kleiner, B. Dalenbäck, and P. Svensson. "Auralization An Overview". In: *Journal of the Audio Engineering Society* 41.11 (1993), pp. 861–875 (cit. on p. 13).

[Kon+09] V. Konz, M. Müller, and A. Scharfstein. "Extracting Expressive Tempo Curves from Music Recordings". In: *Proceedings of the International Conference on Acoustics*. 2009 (cit. on p. 35).

[Käh+04] K. Kähäri, G. Zachau, M. Eklöf, and C. Möller. "The influence of music and stress on musicians' hearing". In: *Journal of Sound and Vibration* 277.3 (2004), pp. 627–631 (cit. on p. 10).

[Lai+11] I. Laird, D. T. Murphy, P. Chapman, and J. Seb. "Development of a Virtual Performance Studio with application of Virtual Acoustic Recording Methods". In: *130th Convention of the Audio Engineering Society*. New York, 2011, p. 12 (cit. on p. 16).

[Lar+16] L. T. Lara, A. M. Pasqual, and M. A. de Mendonça Vecci. "The effects of the excitation source directivity on some room acoustic descriptors obtained from impulse response measurements". In: *Proceedings of the 22nd International Congress on Acoustics*. September. Buenos Aires, 2016 (cit. on p. 66).

[LC67] D. Luce and M. Clark Jr. "Physical Correlates of Brass-Instrument Tones". In: *Journal of the Acoustical Society of America* 42.6 (1967), pp. 1232–1243 (cit. on p. 38).

[Len06] T. Lentz. "Dynamic Cross-Talk Cancellation for Binaural Synthesis in Virtual Reality Environments". In: *Journal of the Audio Engineering Society* 54.4 (2006), pp. 283–294 (cit. on p. 14).

[Ler08] A. Lerch. "Software-Based Extraction of Objective Parameters from Music Performances". PhD thesis. TU Berlin, 2008, p. 178 (cit. on p. 19).

[Ler12] A. Lerch. *Music Performance Analysis*. John Wiley & Sons, Inc., 2012 (cit. on p. 18).

[Lim+13] H. Lim, Y. S. Kim, Y. H. Kim, and J. Y. Jeon. "Preferred acoustical conditions for musicians on stage with orchestra shell in multi-purpose halls". In: *Proceedings of the International Symposium on Room Acoustics*. Toronto, 2013 (cit. on p. 58).

[Lok13] T. Lokki. "Throw Away that Standard and Listen: Your Two Ears Work Better". In: *Proceedings of the International Symposium on Room Acoustics*. Toronto, 2013 (cit. on p. 9).

[Lok14] T. Lokki. "Tasting music like wine: Sensory evaluation of concert halls". In: *Physics Today* 67.1 (2014), pp. 27–32 (cit. on p. 61).

[LT07] O. Lartillot and P. Toiviainen. "A Matlab Toolbox for Musical Feature Extraction from Audio". In: *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-07)*. Bordeaux, 2007 (cit. on pp. 19, 31–34).

[Luc75] D. A. Luce. "Dynamic Spectrum Changes of Orchestral Instruments". In: *Journal of the Audio Engineering Society* 23.7 (1975), pp. 565–568 (cit. on pp. 38, 120).

[Lê+08] S. Lê, J. Josse, and F. Mazet. "Package FactoMineR". In: *Journal of Statistical Software* 25.1 (2008), pp. 1–18. arXiv: `arXiv:1504.04543` (cit. on p. 69).

[Mar+78] A. H. Marshall, D. Gottlob, and H. Alrutz. "Acoustical conditions preferred for ensemble". In: *The Journal of the Acoustical Society of America* 64.5 (1978), pp. 1437–1442 (cit. on pp. 10, 15).

[McK05] C. McKay. “Automatic Music Classification and Similarity Analysis”. In: *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*. 2005 (cit. on p. 128).

[Mey95] J. Meyer. “Influence of communication on stage on the musical quality”. In: *Proceedings of 15th International Congress on Acoustics*. Trondheim, 1995, pp. 573–576 (cit. on p. 10).

[MF09] C. Mckay and I. Fujinaga. “jMIR: Tools for automatic music classification”. In: *Proceedings of the International Computer Music Conference (ICMC ’09)*. 2009, pp. 65–68 (cit. on p. 19).

[MM13] J. V. Mourik and D. T. Murphy. “Geometric and wave-based acoustic modelling using Blender”. In: *AES 49th International Conference: Audio for Games*. 2013, pp. 1–9 (cit. on p. 14).

[MM85] A. H. Marshall and J. Meyer. “The directivity and auditory impressions of singers”. In: *Acta Acustica united with Acustica* 58 (1985), pp. 130–140 (cit. on pp. 12, 15).

[Mül+09] M. Müller, V. Konz, A. Scharfstein, S. Ewert, and M. Clausen. “Towards Automated Extraction of Tempo Parameters”. In: *In Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 2009 (cit. on p. 35).

[Nak84] I. Nakayama. “Preferred Time Delay of a Single Reflection for Performers”. In: *Acta Acustica united with Acustica* 54 (1984), pp. 217–221 (cit. on p. 10).

[Nos+00] D. Noson, S. Sato, H. Sakai, and Y. Ando. “Singer Responses To Sound Fields With a Simulated Reflection”. In: *Journal of Sound and Vibration* 232.1 (2000), pp. 39–51 (cit. on p. 12).

[Nos+02] D. Noson, S. Sato, H. Sakai, and Y. Ando. “Melisma singing and preferred stage acoustics for singers”. In: *Journal of Sound and Vibration* 258.3 (2002), pp. 473–485 (cit. on p. 12).

[OB89] A. M. Ondet and J. L. Barbry. “Modeling of sound propagation in fitted workshops using ray tracing”. In: *The Journal of the Acoustical Society of America* 85.2 (1989), pp. 787–796 (cit. on p. 14).

[Pan+17] L. Panton, D. Holloway, D. Cabrera, and L. Miranda. “Stage Acoustics in Eight Australian Concert Halls: Acoustic Conditions in Relation to Subjective Assessments by a Touring Chamber Orchestra”. In: *Acoustics Australia* (2017), p. 15 (cit. on p. 9).

[Pee04] G. Peeters. *A large set of audio features for sound description (similarity and classification) in the CUIDADO project*. Tech. rep. 2004. arXiv: `0611061v2 [arXiv:quant-ph]` (cit. on p. 19).

[PK97] A. Pietrzyk and M. Kleiner. “The Application of the Finite Element Method to the prediction of Sound Fields of Small Rooms at Low Frequencies”. In: *102th Convention of the Audio Engineering Society*. Gothenburg, 1997 (cit. on p. 14).

[PL10] Jukka Patynen and Tapio Lokki. “Directivities of symphony orchestra instruments”. In: *Acta Acustica united with Acustica* 96.1 (2010), pp. 138–167 (cit. on p. 127).

[PL16] J. Pätynen and T. Lokki. “Concert halls with strong and lateral sound increase the emotional impact of orchestra music”. In: *The Journal of the Acoustical Society of America* 139.3 (2016), pp. 1214–1224 (cit. on pp. 27, 123).

[Pol15] M. Pollow. “Directivity Patterns for Room Acoustical Measurements and Simulations”. PhD thesis. RWTH Aachen University, 2015, p. 176 (cit. on p. 54).

[Pul01] V. Pulkki. “Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning”. In: *J. Audio Eng. Soc.* 49.9 (2001), pp. 753–767 (cit. on p. 28).

[Pul07] V. Pulkki. “Spatial sound reproduction with directional audio coding”. In: *J. Audio Eng. Soc* 55.6 (2007), pp. 503–516 (cit. on p. 13).

[Pul97] V. Pulkki. “Virtual Sound Source Positioning Using Vector Base Amplitude Panning”. In: *Journal of the Audio Engineering Society* 45.6 (1997), pp. 456–466 (cit. on pp. 14, 27).

[Pät+13] J. Pätynen, S. Tervo, and T. Lokki. “Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses.” In: *The Journal of the Acoustical Society of America* 133.2 (2013), pp. 842–57 (cit. on p. 41).

[Pät11] J. Pätynen. “A virtual symphony orchestra for studies on concert hall acoustics”. PhD thesis. Aalto University, 2011, p. 109 (cit. on p. 9).

[Raf03] B. Rafaely. *Plane-wave decomposition of the sound field on a sphere by spherical convolution*. Tech. rep. University of Southampton, 2003 (cit. on p. 13).

[Ram+14] J. Ramo, V. Välimäki, and B. Bank. “High-Precision Parallel Graphic Equalizer”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.12 (2014), pp. 1894 –1904 (cit. on p. 50).

[Rot95] J. Rothstein. *MIDI: A Comprehensive Introduction*. A-R Editions, Inc., 1995, p. 263 (cit. on p. 18).

[Roy+91] J. D. Royster, L. H. Royster, and M. C. Killion. “Sound exposures and hearing thresholds of symphony orchestra musicians”. In: *Journal of the Acoustical Society of America* 89.6 (1991), pp. 2793–2803 (cit. on p. 10).

[Rus+13] F. A. Russo, A. Behar, M. Chasin, and S. Mosher. “Noise exposure and hearing loss in classical orchestra musicians”. In: *International Journal of Industrial Ergonomics* 43.6 (2013), pp. 474–478 (cit. on p. 10).

[Sat+00] S. Sato, Y. Ando, and S. Ota. “Subjective Preference of Cellists for the Delay Time of a Single Reflection in a Performance”. In: *Journal of Sound and Vibration* 232.1 (2000), pp. 27–37 (cit. on p. 10).

[Sav+99] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. “Creating Interactive Virtual Acoustic Environments”. In: *Journal of the Audio Engineering Society* 47.9 (1999), pp. 675–705 (cit. on p. 14).

[Sav10] L. Savioja. “Real-time 3D finite-difference time-domain simulation of low- and mid-frequency room acoustics”. In: *Proceedings of 13th International Conference on Digital Audio Effects (DAFx-10)*. Graz, 2010 (cit. on p. 14).

[SB14] J. Schlüter and S. Böck. “Improved Musical Onset Detection With Convolutional Neural Networks”. In: *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*. 2014 (cit. on p. 19).

[Sea14] C. E. Seashore. “The tonoscope”. In: *The Psychological Monographs* 16.3 (1914), pp. 1–12 (cit. on p. 18).

[Sea37] H. G. Seashore. “An objective analysis of artistic singing”. In: *University of Iowa studies in the psychology of music: Vol. IV. Objective analysis of musical performance*. Ed. by C. E. Seashore. Iowa City: University of Iowa, 1937, pp. 12–157 (cit. on pp. 17, 86).

[Sea38] C. E. Seashore. *Psychology of Music*. New York: Dover Publications, 1938 (cit. on p. 18).

[SK15] Z. Schärer-Kalkandjiev. “The Influence of Room Acoustics on Solo Music Performances. An Empirical Investigation”. PhD thesis. TU Berlin, 2015, p. 227 (cit. on pp. 16, 17, 19–21, 35, 55, 100, 109, 111, 126).

[SKL09] Z. Schärer-Kalkandjiev and A. Lindau. “Evaluation of Equalization Methods for Binaural Signals”. In: *126th Convention of the Audio Engineering Society*. Munich, 2009 (cit. on p. 16).

[SKW13a] Z. Schärer-Kalkandjiev and S. Weinzierl. “Room acoustics viewed from the stage : Solo performers ’ adjustments to the acoustical environment”. In: *Proceedings of the International Symposium on Room Acoustics*. Toronto, 2013 (cit. on p. 16).

[SKW13b] Z. Schärer-Kalkandjiev and S. Weinzierl. "The influence of room acoustics on solo music performance: An empirical case study". In: *Acta Acustica united with Acustica* 99.3 (2013), pp. 433–441 (cit. on p. 20).

[SKW15] Z. Schärer-Kalkandjiev and S. Weinzierl. "The Influence of Room Acoustics on Solo Music Performance: An Experimental Study." In: *Psychomusicology: Music, Mind, and Brain* 25.3 (2015), pp. 195–207 (cit. on p. 21).

[Slo00] J. A. Sloboda. "Individual differences in music performance". In: *Trends in Cognitive Sciences* 4.10 (2000), pp. 397–403 (cit. on p. 17).

[SS15] L. Savioja and U. P. Svensson. "Overview of geometrical room acoustic modeling techniques". In: *The Journal of the Acoustical Society of America* 138.2 (2015), pp. 708–730 (cit. on p. 14).

[TE03] R. Turetsky and D. Ellis. "Ground-Truth Transcriptions of Real Music from Force-Aligned MIDI Syntheses". In: *Proceedings of the 4th International Symposium on Music Information Retrieval (ISMIR)*. 2003, pp. 135–141 (cit. on pp. 31, 34, 35).

[TE16] P. Toiviainen and T. Eerola. *MIDI Toolbox 1.1*. 2016 (cit. on p. 30).

[Ter+13] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki. "Spatial decomposition method for room impulse responses". In: *Journal of the Audio Engineering Society* 61.1/2 (2013), pp. 17–28 (cit. on pp. 13, 25, 38, 51).

[Ter+14] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki. "Preferences of critical listening environments among sound engineers". In: *Journal of the Audio Engineering Society* 62.5 (2014), pp. 300–314 (cit. on p. 27).

[Ter+15a] S. Tervo, J. Pätynen, N. Kaplanis, et al. "Spatial Analysis and Synthesis of Car Audio System and Car-Cabin Acoustics with a Compact Microphone Array". In: *Journal of the Audio Engineering Society* 63.11 (2015), pp. 914–925 (cit. on pp. 27, 51).

[Ter+15b] S. Tervo, J. Saarelma, J. Pätynen, I. Huhtakallio, and P. Laukkanen. "Spatial Analysis of the Acoustics of Rock Clubs and Nightclubs". In: *Proceedings of the Institute of Acoustics*. Vol. 37. 2015, pp. 551–558 (cit. on p. 27).

[TS30] J. Tiffin and C. E. Seashore. "The Iowa Piano Camera". In: *Science* 72.1858 (1930), pp. 146–147 (cit. on p. 18).

[Uen+01] K. Ueno, K. Yasuda, H. Tachibana, and T. Ono. "Sound field simulation for stage acoustics using 6-channel system." In: *Acoustical Science and Technology* 22.4 (2001), pp. 307–309 (cit. on pp. 15, 20).

[Uen+05] K. Ueno, T. Kanamori, and H. Tachibana. "Experimental study on stage acoustics for ensemble performance in chamber music". In: *Acoustical Science and Technology* 26.4 (2005), pp. 345–352 (cit. on pp. 11, 12, 60).

[Uen+10] K. Ueno, K. Kato, and K. Kawai. "Effect of room acoustics on musicians' performance. part I: Experimental investigation with a conceptual model". In: *Acta Acustica united with Acustica* 96 (2010), pp. 505–515 (cit. on pp. 17, 19, 20, 111, 126).

[Uni15] International Telecommunication Union. *ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level*. 2015 (cit. on p. 31).

[Uni16] European Broadcast Union. *Tech 3342 - Loudness Range: A measure to supplement EBU R 128 loudness normalization*. 2016 (cit. on p. 32).

[Uni97] International Telecommunication Union. *ITU-R BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. 1997 (cit. on p. 47).

[UT03] K. Ueno and H. Tachibana. “Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system”. In: *Acoustical Science and Technology* 24.3 (2003), pp. 130–138 (cit. on pp. 11, 60, 66).

[VR81] J. Vos and R. Rasch. “The Perceptual Onset of Musical Tones.” In: *Perception & psychophysics* 29.4 (1981), pp. 323–335 (cit. on p. 35).

[WB11] C. J. Webb and S. Bilbao. “Computing room acoustics with CUDA - 3D FDTD schemes with boundary losses and viscosity”. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2011, pp. 317–320 (cit. on p. 14).

[Wen+12] R. H. C. Wenmaekers, C. C. J. M. Hak, and L. C. J. van Luxemburg. “On Measurements of Stage Acoustic Parameters: Time Interval Limits and various Source–Receiver Distances”. In: *Acta Acustica united with Acustica* 98.5 (2012), pp. 776–789 (cit. on p. 66).

[WG04] G. Widmer and W. Goebl. “Computational Models of Expressive Music Performance : The State of the Art”. In: *Journal of New Music Research* 33.3 (2004), pp. 203–216 (cit. on p. 128).

[WS04] F. Wickelmaier and C. Schmid. “A Matlab function to estimate choice model parameters from paired-comparison data.” In: *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* 36.1 (2004), pp. 29–40 (cit. on pp. 59, 64).

List of Figures

List of Tables

List of Publications

Parts of this thesis and complementary projects have been published in scientific articles. The following articles were written by the current author (in chronological order):

1. **Amengual Garí, S. V.**; Kob, M.: "Perceptual Evaluation of Focused Sources in a Concert Hall". In *Fortschritte der Akustik - DAGA 2015. 41. Deutsche Jahrestagung für Akustik*, pp. 502-505, Nürnberg, Germany, 2015.

2. **Amengual Garí, S. V.**; Lachenmayr, W.; Kob, M.: "Study on the influence of room acoustics on organ playing using room enhancement". In *Proceedings of the Third Vienna Talk on Music Acoustics*, pp. 252-258, Vienna, Austria, 2015.

3. **Amengual Garí, S. V.**; Lachenmayr, W.; Kisić, D.; Kob, M.: "Preliminary results of the effect of reverberation and sound delay on organ playing". In Proceedings of FAMA der DEGA 2015: Seminar on Music Acoustics of the German Acoustical Society, pp. 64-67, Hamburg, Germany, 2015.

4. **Amengual Garí, S. V.**; Lokki, T.; Kob, M.: "Virtual acoustics tools applied to the study of music performance: an overview". In Proceedings of FAMA der DEGA 2015: Seminar on Music Acoustics of the German Acoustical Society, Hamburg, Germany, pp. 13-16, 2015.

5. **Amengual Garí, S. V.**; Eddy, D.; Kob, M.; Lokki, T.: "Real-time auralization of room acoustics for the study of live music performance". In *Fortschritte der Akustik - DAGA 2016. 42. Deutsche Jahrestagung für Akustik*, Aachen, Germany, 2016.

6. **Amengual Garí, S. V.**; Lokki, T.; Kob, M.: "Live performance adjustments of solo trumpet players due to acoustics". In *Proceedings of the International Symposium of Music and Room Acoustics 2016*, La Plata, Argentina, 2016.

7. **Amengual Garí, S. V.**; Pätynen, J.; Lokki, T.: "Physical and perceptual comparison of real and focused sound sources in a concert hall". *Journal of the Audio Engineering Society*, vol. 64 (12), pp. 1014-1025, December 2016.

8. **Amengual Garí, S. V.**; Kob, M: "Investigating the impact of a music stand on stage using spatial impulse responses". *142nd Convention of the Audio Engineering Society*, Berlin, May 2017.

9. **Amengual Garí, S. V.**; Lachenmayr, W.; Mommertz, E.: "Spatial analysis and auralization of room acoustics using a tetrahedral microphone", *J. Acoustical Society of America*, vol. 141 (4), pp. EL369-EL374.

10. **Amengual Garí, S. V.**; Kob, M.; Lokki, T.: "Investigations on stage acoustic preferences of solo trumpet players using virtual acoustics". In *Proceedings of the 14th Sound and Music Computing Conference*, Espoo, July 2017.

The current author contributed to the following articles:

1. Kisić, D.; **Amengual Garí, S. V.**; Hadjakos, A.; Kob, M.: "Extraction and evaluation of temporal musical features from MIDI recordings of organ music". In *Fortschritte der Akustik - DAGA 2016. 42. Deutsche Jahrestagung für Akustik*, Aachen, Germany, 2016.

2. Kob, M.; **Amengual Garí, S. V.**; Sahin, B.; Hadjakos, A.; Saulich, M.: "Online-Tool for interactive sound analysis of orchestra instruments". In *Fortschritte der Akustik - DAGA 2016. 42. Deutsche Jahrestagung für Akustik*, Aachen, Germany, 2016.

3. Kob, M.; **Amengual Garí**, S. V.; Hadjakos, A; Sahin, B.; Saulich, M.; Storjak, I.: "AMISE: Acoustics and Musical Instruments Sound Explorer". In *Proceedings of the 29th Tonmeistertagung - VDT International Conference*, Cologne, Germany, pp. 217-220, 2016.

4. Sahin, B.; **Amengual Garí, S.V.**; Kob, M.: "Investigating listeners' preferences in Detmold Concert Hall by comparing sensory evaluation and objective measurements". In *Fortschritte der Akustik - DAGA 2017. 43 Deutsche Jahrestagung für Akustik*, Kiel, Germany, 2017.

Curriculum Vitae

Personal information

Name:	Sebastià Vicenç Amengual Garí
Date of birth:	5th December 1990
Place of Birth:	Lloret de Vistalegre, Illes Balears (Spain)

Education

2014 – 2017	Ph.D. in Music Acoustics at Hochschule für Musik Detmold (Germany).
2011 – 2014	Degree in Telecommunication Engineering (M.Sc.) at the Polytechnic University of Catalonia (UPC) - BarcelonaTech (Spain). M.Sc. Thesis: "*Alternative Methods to Generate a First-Order Directional Microphone*", completed at the Norwegian University of Science and Technology - NTNU (Norway).
2008 – 2011	Degree in Technical Telecommunication Engineering, with minor in Sound and Image (B.Sc.) at the Polytechnic University of Catalonia (UPC) - BarcelonaTech (Spain). B.Sc. Thesis: "*Implementation of an aircraft noise source and propagation model in Matlab*"
2006 – 2008	High School Diploma (Batxillerat) at IES Madina Mayurqa (Palma, Illes Balears).
2002 – 2006	Secondary School Diploma (ESO) at IES Sineu (Sineu, Illes Balears).
1996 – 2002	Primary School at CEIP Antònia Alzina (Lloret, Illes Balears).

Work Experience

Aug. 2014 – Dec. 2017	Early Stage Researcher of the BATWOMAN Project (Marie Curie Action Project no. 605867) at the Hochschule für Musik Detmold (Germany).
Oct. 2010 – Dec. 2011	Research Intern at the Laboratory of Acoustic and Mechanical Engineering (LEAM) of the Polytechnic University of Catalonia - BarcelonaTech (Spain).

Declaration

I declare that the work included in this thesis is my own work, completed solely and only with the help of the included references, except stated otherwise in the text. This thesis has not been submitted for any other academic degree or professional qualification.

Detmold, September 2017

Sebastià Vicenç Amengual Garí